

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

ТЕОРІЯ СИГНАЛІВ

*Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського як навчальний посібник
для студентів, які навчаються за спеціальністю 153 "Мікро- та наносистемна
техніка"*

Київ
КПІ ім. Ігоря Сікорського
2019

Теорія сигналів [Електронний ресурс] : навч. посіб. для студ. спеціальності 153 «Мікро- та наносистемна техніка» / КПІ ім. Ігоря Сікорського ; уклад.: А.О. Попов. – Електронні текстові данні (1 файл: 7399 Кбайт). – Київ : КПІ ім. Ігоря Сікорського, 2019. – 268 с.

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського (протокол № 10 від 20.06.2019 р.) за поданням Вченої ради Факультету електроніки (протокол № 05/2019 від 23.05.2019 р.)

Електронне мережне навчальне видання

ТЕОРІЯ СИГНАЛІВ

Укладачі: *Попов Антон Олександрович, канд. техн. наук, доц.*

Відповідальний редактор *Тимофєєв Володимир Іванович, д-р техн. наук, проф.*

Рецензентка: *Гармаш Оксана Вікторівна, к.т.н., доц.*

У навчальному посібнику викладено основний теоретичний матеріал дисципліни «Теорія сигналів». До посібника увійшли розділи з опису видів та властивостей сигналів, властивостей та способів опису дискретних систем, основні відомості з лінійної алгебри, необхідні для розуміння загальних підходів до аналізу сигналів. Також в посібник містить матеріали з спектрального та спектрально-часового аналізу, вейвлет-аналізу, перетворення Уолша, перетворення Хартлі, перетворення Гілберта, опис методів нелінійного аналізу сигналів та основні відомості з аналізу двовимірних сигналів.

The tutorial outlines the basic theoretical material of the discipline "Theory of Signals". The manual describes the types and properties of signals, properties and methods for describing discrete systems, basic information on linear algebra, needed to understand common approaches to signal analysis. The manual also contains materials on spectral and time-frequency analysis, wavelet analysis, Walsh transform, Hartley transform, Gilbert transform, description of methods of nonlinear signal analysis and basic information on two-dimensional signal analysis.

© КПІ ім. Ігоря Сікорського, 2019

ЗМІСТ

Вступ.....	5
Мета і задачі аналізу сигналів	5
1. Сигнали	10
1.1. Загальна характеристика сигналів	10
1.2. Класифікація та параметри сигналів	11
1.3. Аналогові, дискретні та цифрові сигнали.....	16
1.4. Аналого-цифрове перетворення сигналів	20
1.5. Цифрові послідовності.....	26
1.6. Простори сигналів	30
1.6.1. Лінійний простір сигналів	31
1.6.2. Метричний простір сигналів	36
1.6.3. Нормований простір.....	38
1.6.4. Гільбертів простір (простір зі скалярним добутком).....	40
1.7. Представлення сигналів в гільбертовому просторі	42
1.8. Теорема ортогонального проектування	50
1.9. Розклади зі збереженням енергії.....	52
1.10. Види та характеристики біомедичних сигналів	52
1.11. Висновки до розділу 1 та рекомендована література	54
1.12. Питання для самоперевірки.....	55
2. Дискретні системи	57
2.1. Загальні відомості про системи обробки сигналів.....	57
2.2. Лінійні дискретні системи	61
2.3. Імпульсна характеристика лінійних дискретних систем.....	66
2.4. Лінійна та кругова згортки	68
2.5. Опис лінійних дискретних систем за допомогою різницевого рівняння	71
2.6. Z- перетворення	74
2.7. Характеристична функція лінійних дискретних систем	76
2.8. Зв'язок між характеристичною функцією та різницевою рівнянням.....	78
2.9. Комплексна частотна характеристика лінійних дискретних систем	79
2.10. Еквівалентність способів описання систем	81
2.11. Висновки до розділу 2 та рекомендована література	82

2.12. Питання для самоперевірки.....	83
3. Спектральний аналіз сигналів за Фур'є	84
3.1. Аналіз неперервних періодичних сигналів.....	84
3.2. Розподіл енергії в спектрі неперервних періодичних сигналів.....	91
3.3. Аналіз неперервних неперіодичних сигналів.....	92
3.4. Властивості перетворення Фур'є неперервних сигналів.....	96
3.5. Розподіл енергії в спектрі неперіодичного сигналу	101
3.6. Аналіз дискретних нескінченних сигналів	102
3.7. Властивості перетворення Фур'є дискретних нескінченних сигналів.....	104
3.8. Зв'язок між спектрами неперервного та дискретного сигналів	107
3.9. Перетворення Фур'є фінітних дискретних сигналів.....	112
3.10. Накладання спектрів	118
3.11. Вибір частоти дискретизації. Теорема Котельникова-Шеннона.....	122
3.12. Дискретизація смугових сигналів.....	127
3.13. Ефект Гіббса.....	132
3.14. Ефект розтікання спектру.....	135
3.15. Аналіз сигналів з віконними функціями.....	139
3.16. Підсумки до розділу 3 та рекомендована література	143
3.17. Питання для самоперевірки.....	144
4. Спектрально-часовий та інші види аналізу сигналів.....	145
4.1. Спектрально-часовий аналіз за Фур'є	145
4.2. Перетворення Хаара	152
4.3. Вейвлет-перетворення.....	156
4.4. Материнські вейвлет-функції.....	163
4.5. Розклад сигналів у фреймах	166
4.6. Діадне вейвлет-перетворення.....	169
4.7. Кореляційний аналіз детермінованих сигналів.....	170
4.8. Перетворення Хартлі.....	174
4.9. Перетворення Уолша-Адамара	176
4.10. Перетворення Гілберта	180
4.11. Підсумки до розділу 4 та рекомендована література	188
4.12. Питання для самоперевірки.....	189
5. Основи Фільтрації неперервних та дискретних сигналів	191

5.1. Шуми та перешкоди в сигналах.....	191
5.2. Типи та характеристики фільтрів	193
5.2. Етапи розробки фільтрів.....	199
5.3. Поняття про апроксимацію фільтрів.....	199
5.4. Цифрові БІХ- та КІХ- фільтри	203
5.5. Основні відомості щодо методів розрахунку коефіцієнтів цифрових фільтрів.....	204
5.5. Побудова структур цифрових фільтрів.....	206
5.6. Застосування фільтрації для багатошвидкісної обробки сигналів.....	210
5.8. Фільтр Вінера.....	218
5.9. Підсумки до розділу 5 та рекомендована література	221
5.8. Питання для самоперевірки.....	222
6. АНАЛІЗ ВИПАДКОВИХ сигналів.....	223
6.1. Випадкові сигнали.....	223
6.2. Основні характеристики випадкових сигналів.....	225
6.2. Класифікація випадкових сигналів.....	228
6.3. Поняття про оцінки характеристик випадкових сигналів.....	230
6.4. Неперервні стаціонарні ергодичні випадкові сигнали	231
6.5. Дискретні стаціонарні ергодичні випадкові сигнали	232
6.6. Спектральний аналіз дискретних стаціонарних ергодичних випадкових сигналів.....	233
Корелограмний метод оцінки спектру	233
Періодограмні методи оцінки СГП	234
6.7. Підсумки до розділу 6 та рекомендована література	237
6.8. Питання для самоперевірки.....	238
7. ОСНОВИ АНАЛІЗУ БАГАТОВИМІРНИХ СИГНАЛІВ.....	239
7.1. Зображення як багатовимірні сигнали	239
7.2. Представлення неперервних та дискретних зображень	241
7.3. Обробка зображень в просторовій області	242
7.4. Підхід до обробки зображень в частотній області.....	251
7.5. Підсумки до розділу 7 та рекомендована література	253
7.6. Питання для самоперевірки.....	253
8. Нелінійний та мультиваріативний аналіз сигналів	255

8.1. Загальна класифікація методів аналізу сигналів.....	255
8.2. Уніваріативний аналіз.....	256
8.3. Біваріативний аналіз.....	262
8.4. Підсумки до розділу 8 та рекомендована література	264
8.5. Питання для самоперевірки.....	265
Використана література.....	266

ВСТУП

Мета і задачі аналізу сигналів

Дані про всі об'єкти навколишнього світу, незалежно від того живий об'єкт чи технічний, можуть бути відомі досліднику лише в результаті вимірювання. За великим рахунком, те, що ми вимірюємо, і є *сигнал*. Інженер і технічна система, яку він будує, мають вимірювати сигнали, отримувати з них дані, а потім, в результаті інтерпретації даних, отримувати інформацію про те, що відбувається з об'єктами спостереження.

Дослідження різних фізичних процесів і явищ, оцінка стану фізичних та біомедичних об'єктів пов'язані з обробкою та аналізом сигналів. Від того, як ми будемо працювати з отриманим сигналом, наскільки якісно, швидко та ефективно обробимо його та проаналізуємо його параметри, як саме проінтерпретуємо параметри і які зробимо висновки – залежить те, наскільки багато інформації про сам об'єкт спостереження ми отримаємо з сигналу, і наскільки достовірною та актуальною буде ця інформація. Тому, як аналізувати сигнали в технічних системах і присвячено даний посібник.

Нехай існує певний об'єкт спостереження, який необхідно дослідити (отримати достовірну інформацію про його стан та поведінку). Ми знаємо або припускаємо, що стан об'єкту змінюється з часом, внаслідок проходження ньому деяких фізичних процесів. Об'єктом може бути, наприклад, датчик певної фізичної величини, що під впливом полів змінює свій опір; антена, яка випромінює в ефір електромагнітне поле; серце людини, робота якого супроводжується виникненням змінної в часі різниці потенціалів на поверхні тіла людини; голосові зв'язки нашого співрозмовника, коливання яких причиняє виникнення звуку – хвилі механічних коливань молекул газів повітря, які ми можемо сприйняти вухами.

Внаслідок проходження фізичних процесів змінюються характеристики об'єкта (опір, температура, колір, вага, напруга, струм, тиск, спектр випромінення та т.і.) Те, наскільки інтенсивно в об'єкті протікають ті чи інші фізичні процеси, проявляється у вигляді числового результату вимірювання фізичної величини. Задача технічної системи, яка розроблена для аналізу стану об'єкта – достовірно сприйняти ці зміни. Для дослідника всі дані про характеристики об'єкту та їх зміну є результатами вимірювання деякого параметра – фізичної величини. Ці результати розглядаються як сигнали, які супроводжують існування об'єкта спостереження.

Зазвичай необхідно не лише реєструвати та вимірювати сигнали, які супроводжують існування об'єкту. Треба із цих сигналів отримати інформацію про стан та поведінку об'єкта, а часто і спрогнозувати його стан у майбутньому або сформулювати деякий керівний вплив на об'єкт, щоб перевести його в потрібний стан. В цьому випадку говорять про обробку та аналіз сигналів. Види аналізу сигналів можна розділити на такі:

1. Описовий аналіз (descriptive analytics) – отримання характеристик (чисельних та якісних) сигналів, що реєструються від об'єкта спостереження.
2. Прогнозування (predictive analytics) – отримання відомостей про те, що станеться з об'єктом у майбутньому на основі його характеристик в теперішньому стані. Додатково можуть залучатися історичні дані про поведінку об'єкту, а також моделі об'єкта. Прогнозування полягає у знаходженні зв'язку між причиною та наслідком, та виконується з використанням статистичних методів, регресії, класифікації.
3. Прескриптивний (наказовий) аналіз (prescriptive analytics) – застосування до існуючих даних про об'єкт методів, які дозволять визначити такі дії об'єкта (або над об'єктом), що призведуть до його переходу в потрібний стан. В результаті такого аналізу виробляються рекомендації до потрібних дій.
4. Когнітивний аналіз (cognitive analytics) – сукупність методів, які намагаються відтворити аналітичні можливості людського мозку для того, щоб робити висновки на основі наявних даних, розпізнавати характерні ситуації використовувати знання про попередню поведінку для того, щоб отримувати ще більше знань про об'єкт чи процес. Для цього використовуються складні методи штучного інтелекту, самонавчання, навчання з підкріпленням.

Необхідно чітко розрізняти поняття сигналу, інформації, даних, повідомлення і знань про об'єкт. Те що ми вимірюємо, носій відомостей про зміну фізичної величини – сигнал. В технічній системі, яка призначена для аналізу та обробки сигналів, сигнал є, як правило, деякою напругою або струмом, що пропорційні зміні вимірюваної величини. Результат вимірювання сигналу – дані; як правило, в цифрових системах аналізу сигналів – це числові значення відліків сигналу в певні моменти часу. Ця послідовність чисел, яку має наша система в результаті вимірювання – єдине, що вона «знає» про сигнал (і про об'єкт спостереження). Тому задача дослідника – забезпечити всі умови для того, щоб всі суттєві властивості початкового сигналу, якій треба було виміряти, були збережені в цих цифрових даних.

Інформація – це результат інтерпретації даних, відомості про певне явище, подію, об'єкт. Дані (символи), які представлені у певній формі – це повідомлення, тобто те, що підлягає передачі, отже інформація передається у вигляді повідомлень. Повідомлення передається з допомогою деякого поля, що є матеріальним носієм повідомлення. Це може бути матеріальне поле (наприклад, людська мова при розмові передається з допомогою змін локального тиску повітря, які розповсюджуються від голосових зв'язок одного співрозмовника до барабанної перетинки слухача), або електромагнітне поле (наприклад, радіохвилі, які використовуються в радіозв'язку). Повідомлення треба сформувати, іноді – закодувати, передати по каналу зв'язку, потім зареєструвати та вилучити з нього дані.

Знання про об'єкт – є результатом інтерпретації отриманої інформації.

Зрозуміло, що один і той самий сигнал може бути корисним або непотрібним залежно від того, чи можемо ми інтерпретувати дані, які він нам несе. Наприклад, якщо ми зареєструємо зашифрований сигнал, то від того, чи відомий нам ключ до шифру, буде залежати, чи зможемо ми дістати з зашифрованого повідомлення корисну інформацію чи ні.

Часто в реальних умовах реєстрації сигналів на вхід системи потрапляє не тільки корисний сигнал від того об'єкту, який ми досліджуємо, а і інші сигнали від сторонніх об'єктів (шуми, перешкоди та завади). Наявність шумів ускладнює аналіз сигналів та подальшу інтерпретацію отриманих даних, оскільки часто дуже складно розрізнити, що в сумарному сигналі є корисною складовою, а що в ньому з'явилося внаслідок зашумлення. Одною з задач обробки сигналів є позбавлення сигналів від небажаних шумових складових за допомогою фільтрації. Але поряд з цим, треба мати на увазі, що шумовий сигнал і його аналіз є дуже корисним в тому випадку, коли ми досліджуємо саме характеристики джерела шуму.

Дослідження сигналів – процес, спрямований на отримання інформації, яку містять сигнали.

Мета дослідження сигналів – формування достовірного уявлення про стан та властивості об'єкту шляхом обробки та аналізу сигналів, що супроводжують його існування.

З дослідженням сигналів пов'язані такі *задачі*:

1. *генерація* сигналів із заданими характеристиками;
2. *передача* сигналів по каналах зв'язку;
3. *реєстрація* сигналів;
4. *вимірювання* сигналів;
5. *кодування* сигналів для передачі з їх допомогою потрібних даних;
6. *стиснення* сигналів;
7. *відображення* сигналів та результатів їх перетворень;
8. *збереження* сигналів і результатів їх перетворень;
9. *обробка* сигналів, зокрема, позбавлення від небажаних складових в сигналі.
Наприклад, фільтрація сигналів для позбавлення від шуму, або розділення каналів при прийманні сигналів мобільного зв'язку;
10. *перетворення* сигналів, зокрема *перетворення формату сигналу*, наприклад, дискретизація сигналу – перетворення неперервного сигналу на послідовність дискретних значень сигналу в певні моменти часу;
11. *розрахунок параметрів* сигналу, які пов'язуються дослідником з корисною інформацією про об'єкт спостереження. Як правило, інформація може бути скрита в кількісних характеристиках: енергії, потужності, амплітуді, фазі, частоті, спектральних співвідношеннях, часових залежностях між різними сигналами та т.і. Загалом, все, що ми можемо поррахувати на основі вимірюного сигналу, може бути корисним параметром для деякої мети аналізу сигналів;

12. *аналіз* сигналів з метою одержання найбільш важливої інформації про об'єкт дослідження. В переважній більшості випадків аналіз сигналів ведеться шляхом *розкладу сигналу* на елементарні складові для їх окремого розгляду або для порівняння властивостей різних сигналів. Класичний приклад – спектральний аналіз за Фур'є, при якому кожний сигнал розкладається на гармонічні синусоїдальні коливання, і потім вивчається вираженість (амплітуда) складових коливань на тих чи інших частотах в даному сигналі. Ми будемо вивчати розклади не лише за Фур'є, а і з використанням інших базисних функцій. Від того, на які складові ми розкладемо сигнал, буде залежати, які саме властивості ми виявимо в ньому;
13. *визначення зв'язку між сигналами* та характеристик одного сигналу на основі характеристик іншого, який з ним зв'язаний. Для цього використовуються методи біваріативного або мультिवаріативного аналізу сигналів, які дозволяють чисельно оцінити силу зв'язку, синхронності та схожості між сигналами;
14. *прогнозування* майбутніх характеристик сигналу на основі знань про його характеристики в минулому;
15. *розпізнавання* образів сигналів, діагностика та прогнозування стану об'єкта.

Галузі застосування сигналів:

– телекомунікації, наприклад, радіо- та мобільний зв'язок, мережа Інтернет, телебачення (позбавлення від луни, ущільнення каналів за технологіями FDMA, CDMA, TDMA, OFDM);

– біомедичні застосування у діагностиці, терапії та моніторингу стану людини (системи аналізу ЕКГ, ЕЕГ, УДЗ, МРТ та ін., системи для зв'язку мозок-комп'ютер, мобільні застосування, безконтактне вимірювання параметрів показників життєдіяльності). В результаті аналізу біомедичних сигналів вирішуються задачі детектування наявності хвороби, прогнозування її перебігу або появи загрозливих станів, попередження захворювань та формування рекомендацій для користувача, щоб захворювання не з'явилося в майбутньому;

– стиснення інформації (наприклад, стиснення рухомого зображення та звуку MPEG, стиснення стаціонарного зображення JPEG, телебачення HDTV, шифрування, покращення якості звуку та зображення);

– системи управління, контролю, прогнозування та попередження;

– розпізнавання та синтез мови;

– військові застосування (безпечний зв'язок, радары);

– дослідження нових матеріалів, техніка спектроскопії та МРТ.

Базовими дисциплінами, на яких має ґрунтуватися вивчення теорії сигналів є лінійна алгебра, аналітична геометрія та математичний аналіз, імовірнісні основи обробки даних та математична статистика, обчислювальна математика та інформатика, метрологія та техніка вимірювань. Але якщо студент опанував

принаймні знаннями з вищої математики в обсязі університетського курсу, то вивчення даної дисципліни не повинно скласти труднощів.

В цій книзі основна увага буде зосереджена на методах обробки та аналізу сигналів, а питання інтерпретації, класифікації та прогнозування відносяться до суміжних дисциплін.

Метою дисципліни «Теорія сигналів» є формування цілісного уявлення про сигнали, математичні основи методів їх дослідження, а також створення підґрунтя для реалізації та застосування методів аналізу сигналів.

1. СИГНАЛИ

1.1. Загальна характеристика сигналів

Термін «сигнал» пішов від латинського *signum* – знак. Одного усталеного визначення поняття «сигнал» наразі не існує. Найкоротше та одне з найбільш влучних означень дав В. Даль: «Сигнал – это вестъ знаком» (рос.) («Сигнал – це звістка за допомогою знаку»). В цьому визначенні відображені основні властивості сигналу: те, що в сигналі міститься певна важлива для дослідника інформація, і те, що сигнал не може існувати сам по собі, а обов'язково повинен бути носій, що переносе енергію сигналу від джерела до адресата. Наразі неможливо передати інформацію не передаючи при цьому енергію, а отже можна сказати, що сигнал – це сукупність повідомлення та його носія. Носієм сигналу є поле: електромагнітне, температурне або поле матеріальних часток. Вимірюючи характеристики такого поля (фізичні величини) ми вимірюємо сигнали. Значення сигналу – це значення фізичної величини, яка вимірюється у часі, просторі чи інших координатах.

Те, що інформація передається не безпосередньо, а з використанням знаків, символів, фізичних величин, що несуть змістовне навантаження, передбачає необхідність обробки, аналізу та інтерпретації тих даних, що будуть в розпорядженні адресата повідомлення.

Сигнал – це фізичний процес, характеристики якого несуть інформацію про певне явище чи стан об'єкта, або передають команди управління та оповіщення. Якщо сигнал спеціально створений для передачі повідомлення, то певний параметр цього процесу (наприклад, величина сигналу, його частота та ін.) змінюється відповідно до тої інформації, яку треба передати.

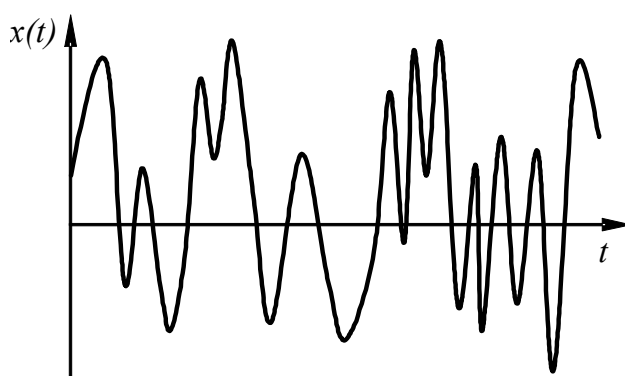


Рисунок 1.1 – Довільний неперервний сигнал $x(t)$, що змінюється в часі

Всі фізичні процеси з точки зору теорії сигналів описуються математичними функціями з тими чи іншими властивостями. Здебільшого це неперервні функції свого аргументу, або числові послідовності (впорядковані набори чисел). На

рисунку 1.1 зображено графік сигналу $x(t)$ – сукупність точок, які відповідають значенню сигналу $x(t)$. Часто в теорії сигналів, якщо це не заважає сприйняттю, поняття «функція» та «сигнал» вживаються як синоніми.

Як правило, характеристики об'єктів змінюються з часом або в просторі. Отже, функції, які будуть описувати зміну величини сигналу – функції часу або просторової координати. Для загальності далі в цій книзі будемо мати на увазі, що всі сигнали, про які іде мова – є функціями часу, якщо інше спеціально не зазначене окремо.

Властивості, які має сигнал, визначають, які саме математичні функції треба використовувати для того, щоб його представити графічно та описати математично. Так, періодичні неперервні сигнали зручно описувати періодичними функціями, а якщо значення сигналу мають випадковий характер, то його зручно описати математично як випадкову функцію. Якщо сигнал дискретний, то його описують як числову послідовність.

Процедуру описання сигналу за допомогою математичної функції можна розглядати як процедуру побудови математичної моделі сигналу. Математична модель сигналу – це створена або обрана дослідником математична функція чи система функцій, яка відображає істотні властивості сигналу, який досліджується. Дослідження моделі сигналу може до певної міри замінити дослідження реального сигналу та дати нову інформацію про об'єкт, з яким сигнал пов'язаний. Моделюванню сигналів присвячений окремий розділ теорії сигналів.

Отже, насправді весь математичний апарат теорії сигналів працює не з реальними сигналами (напругами, тисками чи інтенсивностями), а з їх математичними моделями – функціями, які описують сигнали.

1.2. Класифікація та параметри сигналів

Оскільки і реальні об'єкти, і сигнали, які відповідають їх характеристикам, мають багато різноманітних параметрів, сигнали можна класифікувати за багатьма ознаками. Від того, які властивості мають сигнали, залежить вид функції, яка використовується для математичного описання сигналу.

За причиною виникнення або за об'єктом, з яким пов'язаний сигнал, їх поділяють на:

- а) сигнали природних систем, тобто фізичних та біологічних об'єктів. Прикладом може бути сигнал електроенцефалограми – запис електричних коливань, що супроводжують роботу мозку, або випромінення зірок, що реєструється інфрачервоними телескопами;
- б) сигнали штучних систем, тобто технічних пристроїв. Наприклад, таким сигналом є сигнал, зареєстрований антеною стільникового телефону.

Залежно від властивостей функцій, що описують неперервні сигнали, їх можна поділити на такі види, які часто зустрічаються на практиці.

1. Періодичні сигнали

Сигнал, значення якого повторюються через однакові проміжки часу, називається **періодичним**. Це сигнали, які описуються періодичними функціями. Періодична функція – яка задовольняє таку умову:

$$x(t) = x(t + nT), \quad \forall t \in \mathbb{R}, \quad n \in N, \quad (1.1)$$

де T – постійна величина, яка називається **періодом** – інтервал часу (в загальному випадку – інтервал зміни аргументу), протягом якого відбувається одне повне повторення значень сигналу. Число n може приймати будь-яке значення з множини цілих чисел. Також періодом сигналу будуть також всі величини, які кратні T .

Періодичний сигнал – це математична абстракція; насправді реальних сигналів виду (1.1) немає і бути не може, оскільки в реальному житті не може існувати точного повторення процесів, особливо якщо мова іде про живі об'єкти. Але якщо сигнал, який аналізується, проявляє періодичні властивості, то описання його періодичною функцією може бути корисним.

Також сигнали бувають неперіодичними (однократними) (рис. 1.2).

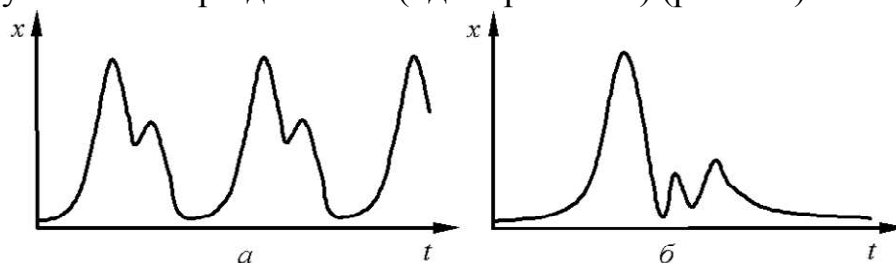


Рисунок 1.2 – а) періодичний сигнал, б) однократний сигнал.

Якщо сигнал змінюється в часі, то величина, обернена до періоду, називається **частотою** періодичного сигналу $f = \frac{1}{T}$ (Гц). В теорії сигналів використовують також поняття **кутової частоти** (циклічної частоти, кутової частоти):

$$\omega = 2\pi f = \frac{2\pi}{T} \text{ (рад/с)}.$$

Введемо позначення для множини періодичних сигналів:

$$S_T = \{x(t): x(t + nT) = x(t), -\infty < t < \infty, n \in N\} \quad (1.2)$$

Окремою підмножиною множини S_T є множина синусоїдальних сигналів. Це сигнали вигляду

$$x(t) = A \sin(\omega t + \varphi).$$

де A – амплітуда сигналу;

ω – частота, рад/с;

φ – початкова фаза сигналу, рад.

2. Сигнали, обмежені за амплітудою

Це сигнали, миттєві значення яких обмежені за амплітудою певним дійсним додатним числом K :

$$S_K = \{x(t) : |x(t)| \leq K, -\infty < t < \infty, K \in \mathbb{R}^+\}.$$

3. Сигнали з скінченною енергією

Існує великий клас сигналів, які називають сигналами зі *скінченною (обмеженою) енергією*. Цей клас сигналів має широке застосування в електроніці, і надалі ми будемо працювати виключно з такими сигналами. Всі сигнали, які нас оточують є сигналами зі скінченною енергією. Множина таких сигналів може бути задана так:

$$S_E = \left\{ x(t) : \int_{-\infty}^{+\infty} |x(t)|^2 dt < E, E \in \mathbb{R}^+ \right\}. \quad (1.3)$$

Про такі сигнали говорять, що їх енергія обмежена величиною E . Фізично інтеграл у формулі (1.3) трактують як енергію, маючи на увазі, що $x(t)$ є напругою, яка падає на опорі в 1 Ом, або струмом, що протікає через такий опір. Функції, які описують сигнали зі скінченною енергією, в літературі часто позначають як $x(t) \in L^2(\mathbb{R})$ і говорять, що вони є інтегрованими за Лебегом з квадратом.

4. Сигнали з обмеженою тривалістю

Це сигнали, значення яких дорівнюють нулю за межами деякого інтервалу аргументу (часу) тривалістю D :

$$S_D = \{x(t) : x(t) = 0, t \notin [0, D]\}. \quad (1.4)$$

Такі сигнали часто називають *фінітними (скінченними) сигналами*. Вони існують лише протягом певного проміжку часу. Поза цим проміжком вони або дорівнюють нулю, або невизначені. Математично такі сигнали описуються функціями з компактним носієм, тобто лише для частини своєї області визначення ця функція не дорівнює нулю.

5. Сигнали з обмеженим спектром

Це сигнали, спектр за Фур'є яких сконцентрований лише на певному проміжку частот, а складових на інших частотах в спектрі сигналу немає. Часто, коли говорять про сигнали з обмеженим спектром, мають на увазі, що спектр лежить на частотах в околі нульової частоти:

$$S_W = \left\{ x(t) : \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt = 0 \quad \forall |\omega| > W \right\}. \quad (1.5)$$

Сигнали з обмеженим спектром – дуже важливий клас сигналів, які відіграють ключову роль в теорії сигналів, оскільки саме для таких сигналів можна наразі розробити дієві методи обробки та аналізу. Більше того, одною з найперших та найважливіших операцій в обробці сигналів зараз є навмисне обмеження його спектру перед дискретизацією за допомогою фільтрації (див. подальші розділи).

Всі сигнали, які підлягають обробці, розділяють на два великі класи за тим, яким математичним апаратом користуються при їх дослідженні і як саме ми розглядаємо сигнал з метою аналізу і які саме параметри і характеристики хочемо отримати в результаті. Виділяють:

Детерміновані сигнали – для яких відоме або можна розрахувати значення у будь-який момент часу в минулому або майбутньому. Такі сигнали описуються однозначними математичними функціями.

Випадкові сигнали – для яких значення в будь-який момент часу є випадковою величиною, що набуває певні значення з деякою імовірністю. Вони описуються імовірнісними характеристиками – математичним сподіванням, дисперсією, густинами або функціями розподілу ймовірностей, моментними функціями та ін.

Переважає більшість даного посібника присвячена дослідженню детермінованих сигналів. Варто зауважити, що як тільки сигнал, який вважається випадковим за своєю природою, виміряний за допомогою деякої системи, зареєстрований та записаний в пам'ять комп'ютера – він може вважатися детермінованим, оскільки його значення відомі нам точно.

Також треба мати на увазі, що те, як ми розглядаємо сигнали, залежить лише від нашого бажання та мети, яку ми перед собою ставимо. Один і той самий сигнал можна розглядати і як випадковий, і як детермінований та характеризувати різними

параметрами. Ці можливості часто використовують при дослідженні біомедичних сигналів.

За кількістю каналів реєстрації сигналу від одного об'єкта сигнали можуть бути *одноканальні* (наприклад, сигнал, зареєстрований антеною радіоприймача) та *багатоканальні* (Рис. 1.3). Для багатоканальних сигналів характерна синхронність сигналів по каналах, тобто одночасність вимірювання чи генерації сигналів. Прикладами таких сигналів може бути електроенцефалограма, напруги керування трьома променями червоного, синього та зеленого кольорів моніторів, сигнали керування градієнтними полями у магніторезонансній томографії.

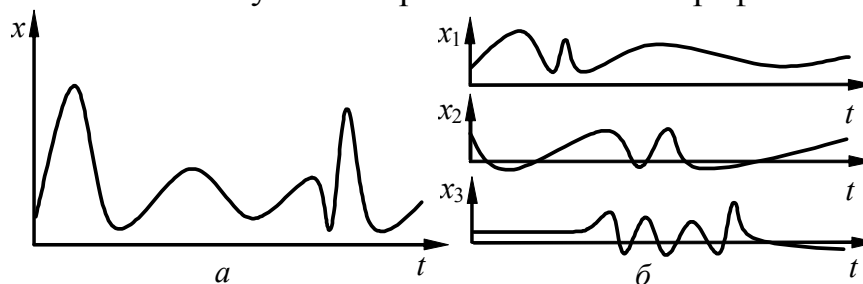


Рисунок 1.3 – а) одноканальний сигнал, б) триканальний сигнал

За кількістю вимірів сигнали поділяють на:

Одновимірні сигнали, коли функція, яка його описує, залежить лише від одної змінної (напруга, струм): $y = x(t)$.

Багатовимірні сигнали (коли можна визначити залежність значення сигналу від кількох змінних): $y = x(t, s, a, z)$. Багатовимірні сигнали існують зазвичай у природних системах. Прикладом таких сигналів може бути розподіл електричного потенціалу по поверхні шкіри людини чи розподіл густини хімічних сполук, температури в біологічних об'єктах, траєкторії руху точок на тілі людини тощо. Але на практиці реєстрація цих сигналів виконується одновимірними паралельними чи послідовними каналами окремо. За допомогою перетворення зазначених сигналів (реконструкції) одержують моделі багатовимірних сигналів і відповідні образи для сприйняття. Прикладом таких зображень є двовимірні томограми, тривимірні образи у медичній інтроскопії.

Також можна класифікувати сигнали:

- за частотним складом (низькочастотні, високочастотні, широкосмугові, вузькосмугові);
- за наявністю модуляції (модульовані або без модуляції);
- за наявністю кодування (кодовані або без кодування);
- за формою (синусоїдальні, гармонічні, імпульсні, довільної форми тощо) (рис. 1.4).

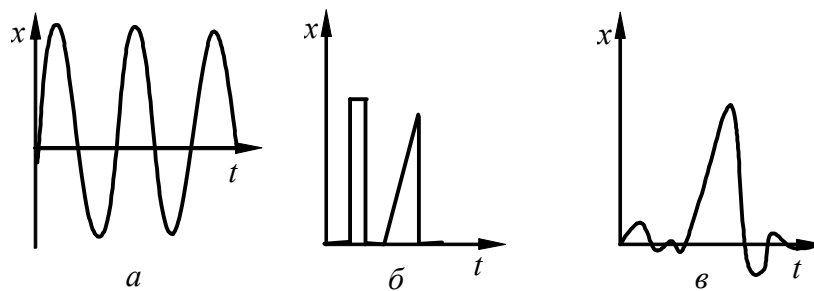


Рисунок 1.4 – Приклади сигналів: а – синусоїдальний сигнал; б – імпульсний сигнал; в – сигнал довільної форми

1.3. Аналогові, дискретні та цифрові сигнали

Важливою характеристикою сигналів є множина значень сигналу та характеристика того, коли саме значення сигналу виміряні або розраховані. З точки зору теорії сигналів ця характеристика впливає на те, якою є область визначення та область значень функції, що описує сигнал. Виділяють чотири класи сигналів за характеристиками математичних функцій, що їх описує:

I. Неперервні у часі та за величиною сигнали.

Такі сигнали ще називають *аналоговими (неперервними) сигналами*. Майже всі сигнали, які існують в реальному світі, є аналоговими, оскільки, як правило, всі фізичні величини змінюються неперервно. Саме аналогові сигнали надходять на вхід систем обробки сигналів з різноманітних сенсорів та антен, і саме вони несуть всю інформацію про об'єкт. Особливістю аналогових (часто їх називають просто неперервними) сигналів є те, що конкретне числове значення сигналу можна визначити в *будь-який* момент часу. Це значення також можуть бути довільними з множини дійсних або комплексних чисел. Математично аналогові сигнали описуються неперервними функціями свого аргументу (рис. 1.5):

$$x(t) \in \mathbb{R}, \quad t \in \mathbb{R}.$$

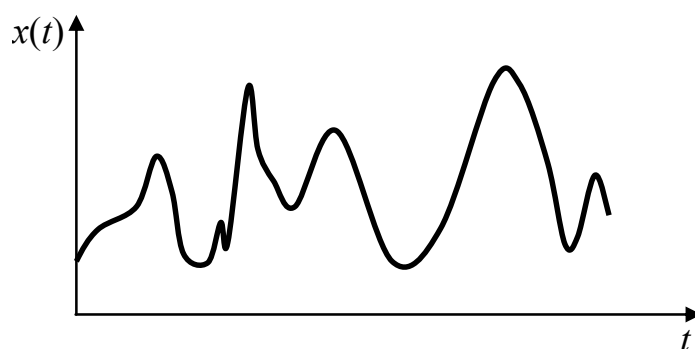


Рисунок 1.5 – Аналоговий сигнал

І значення функції, і значення аргументу є дійсними числами.

II. Сигнали, неперервні в часі та квантовані за величиною.

Такі сигнали визначені на всій числовій вісі, тобто як і для попереднього класу, в будь-який довільний момент часу можна отримати конкретне значення сигналу. Але відмінність полягає в тому, що величина сигналу не може приймати довільне значення, а може бути рівною лише якомусь одному значенню з деякої дискретної множини чисел $\{x_n\}$ (рис. 1.6). Математично такі сигнали описуються неперервними кусково-постійними функціями з розривами першого роду.

$$x(t) = x_n, \quad x_n \in \mathbb{R}, \quad t \in \mathbb{R}, \quad n \in \mathbb{Z}.$$

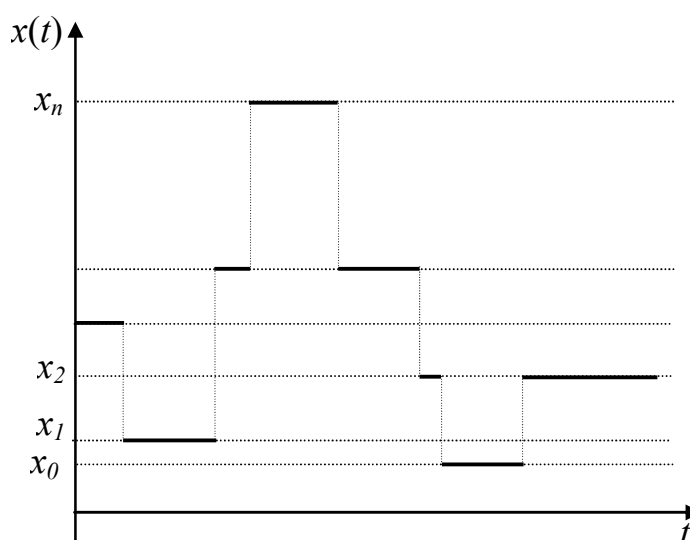


Рисунок 1.6 – Квантований за величиною неперервний сигнал

В переважній більшості випадків значення квантованого сигналу задається у вигляді цілої кількості елементарних кроків квантування (квантів) Δ_x : $x_k = k\Delta_x$, $k \in \mathbb{Z}$. Тоді говорять про рівномірне квантування сигналу (рис. 1.7).

Крок квантування – це різниця між двома сусідніми допустимими значеннями дискретного сигналу.

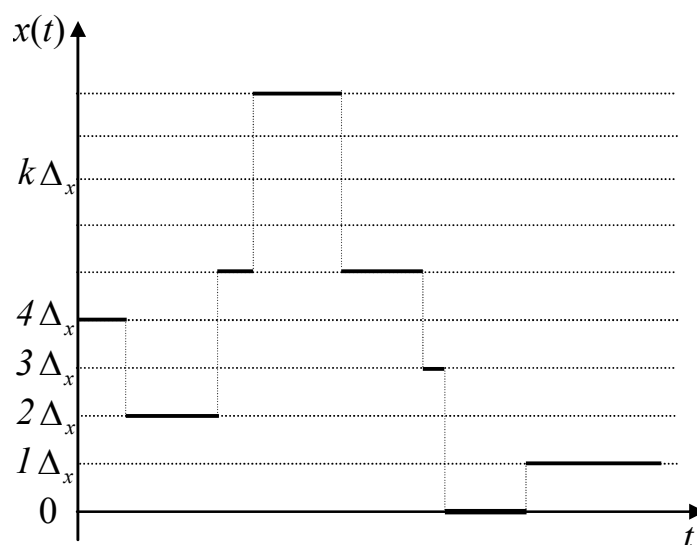


Рисунок 1.7 – Квантований за величиною неперервний сигнал, рівномірне квантування

Сигнали, квантовані за величиною та неперервні в часі, є вихідними сигналами цифро-аналогового перетворювача (ЦАП). ЦАП є генератором постійної напруги, який керується цифровим способом. На вхід ЦАП для кожного відліку дискретного сигналу подається двійковий код значення величини сигналу в певний момент часу (яка виражена цілим числом кроків квантування), а на виході цей пристрій формує неперервний кусково-постійний сигнал. Значення цієї постійної напруги підтримується незмінним протягом одного кроку дискретизації. Після надходження на вхід ЦАП коду для наступного значення напруги, вихідна напруга змінюється, і все повторюється.

III. Сигнали, дискретні в часі та неперервні за величиною.

Сигнали, які дискретні в часовій області та неперервні за величиною, називають *дискретними сигналами*. Їх особливість в тому, що вони існують лише в певні визначені моменти часу (моменти відліків t_n , рис. 1.8). Але величина сигналу може приймати довільне значення з множини дійсних чисел. Між моментами часу t_n , коли заданий дискретний сигнал, такий дискретний сигнал не існує, він в цей час не визначений. Не можна казати, що він дорівнює нулю між відліками, просто дискретний сигнал там не заданий. В ці проміжки часу сигнал не існує і не може бути відновлений ніякими математичними методами без додаткових припущень. Математична функція, яка описує дискретний сигнал, визначена не на всій числовій вісі, а задана лише для певних дискретних значень свого аргументу, тобто є числовою послідовністю:

$$x(t) = x(t_n), \quad x(t_n) \in \mathbb{R}, \quad t_n \in \mathbb{R}, \quad n \in \mathbb{Z}.$$

Якщо дискретний сигнал заданий в моменти часу, які віддалені один від одного на *однаковий* інтервал часу Δ_t , то говорять про *рівномірну* дискретизацію. В цьому випадку $t_n = n\Delta_t$, $n \in \mathbb{Z}$.

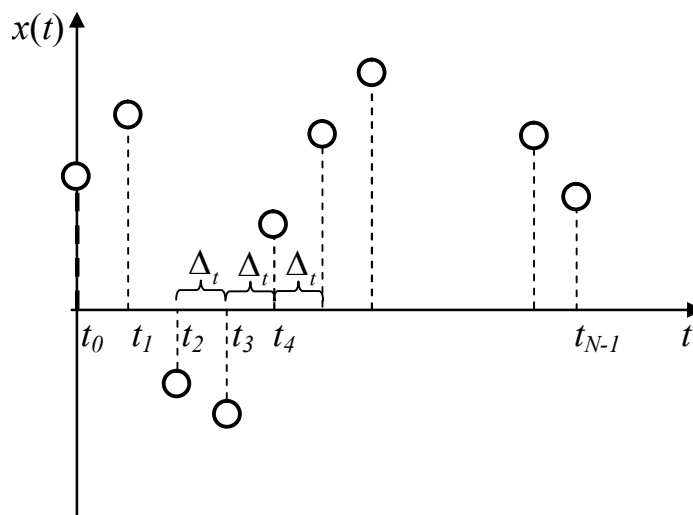


Рисунок 1.8 – Дискретний сигнал з рівномірною дискретизацією

IV. Сигнали, дискретні в часі та квантовані за величиною.

Такі сигнали називають *цифровими сигналами* (рис. 1.9). Цифрові сигнали задані не лише своїми значеннями тільки в певні моменти часу t_i (є дискретними у часі); в ці моменти часу значення сигналу можуть бути лише з деякої дискретної множини (квантовані за величиною):

$$x(t) = x_k(t_n), \quad x_k = k\Delta_x, \quad t_n = n\Delta_t, \quad n \in \mathbb{Z}.$$

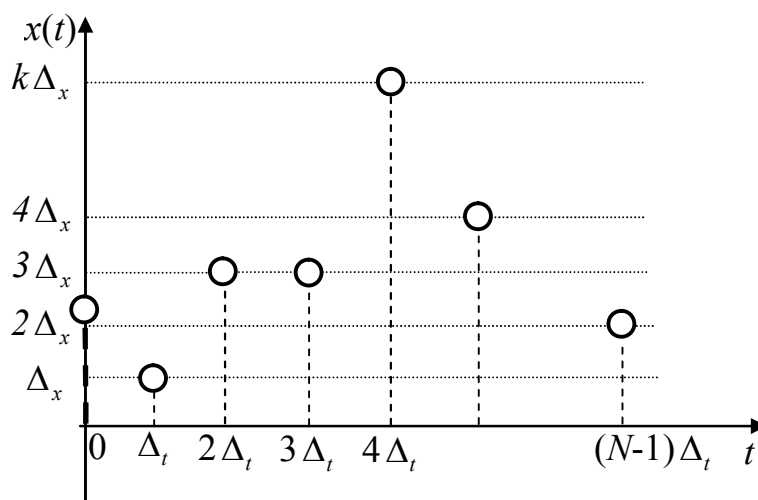


Рисунок 1.9 – Цифровий сигнал

Представлення всіх сигналів у комп'ютерних системах є цифровим. Значення сигналу в дискретні моменти часу представляється у вигляді двійкового коду. Ми розглянемо питання оцифровки сигналів в подальших розділах.

Далі в цьому посібнику ми будемо мати справу лише з неперервними та дискретними сигналами.

1.4. Аналого-цифрове перетворення сигналів

Всі системи цифрової обробки працюють з сигналами, які представлені у вигляді послідовностей чисел – цифровими сигналами. Але оскільки всі реальні сигнали, що зустрічаються в практичній діяльності, є неперервними, то першим етапом будь-якої обробки є перетворення сигналу з аналогового вигляду у цифровий. Це виконується спеціальними пристроями – аналого-цифровими перетворювачами, в кілька етапів. Загальною метою, яку треба досягти при аналого-цифровому перетворенні, є необхідність якомога повного відображення всіх властивостей початкового неперервного сигналу у результуючій послідовності чисел. Якщо це буде забезпечено, то на основі аналізу цифрового сигналу дослідник зможе отримати інформацію про початковий неперервний сигнал (а отже і про об'єкт, з яким даний сигнал пов'язаний).

1. Дискретизація.

Дискретизація неперервного сигналу – перетворення початкового неперервного сигналу $x(t)$ у послідовність миттєвих значень цього сигналу $x(t_n)$, виміряних у фіксовані моменти часу t_n . Послідовність отриманих в результаті дискретизації чисел з математичної точки зору є числовим рядом, тобто впорядкованим набором чисел.

Числа, які утворюють послідовність, називаються *відліками* сигналу (samples, отсчеты). Дискретизація неперервного сигналу може бути рівномірною (циклічною, періодичною), адаптивною або випадковою.

При *рівномірній* дискретизації відліки сигналу беруться через однакові проміжки часу T_s , які визначаються заздалегідь та не залежать від властивостей того сигналу, що поступає на вхід АЦП.

Крок дискретизації (період дискретизації, інтервал дискретизації) T_s – це інтервал часу між сусідніми відліками.

При *адаптивній* дискретизації крок, через який вимірюються значення відліків визначається швидкістю зміни сигналу. На ділянках швидких змін крок дискретизації обирають меншим, ніж на ділянках, де зміни сигналу повільні. При цьому задаються припустимою похибкою між реальним сигналом та його

дискретним значенням. І щойно сигнал відхиляється від прогнозованого значення на величину, більшу похибки, робиться наступне вимірювання відліку сигналу.

Найбільш розповсюдженою є рівномірна дискретизація, при якій послідовність дискретних відліків отримується з неперервного сигналу за формулою:

$$x[n] = x(t_n) = x(nT_s), \quad -\infty < n < +\infty.$$

У випадку рівномірної дискретизації величина обернена періоду,

$$F_s = \frac{1}{T_s}, \text{ Гц}$$

називається *частотою дискретизації*.

Частота дискретизації – кількість періодів дискретизації за одиницю часу. Наприклад, якщо частота дискретизації дорівнює 512 Гц, то за одну секунду буде отримано 512 відліків дискретного сигналу, а інтервал часу між двома сусідніми відліками буде рівний $T_s = \frac{1}{F_s} = \frac{1}{512} \approx 0.002 = 2 \text{ мс}$. Якщо частота дискретизації

вимірюється в рад/с, то вона показує, скільки дискретних значень сигналу буде виміряно за один період сигналу.

В результаті дискретизації ми отримуємо *дискретний сигнал* – послідовність відліків, взятих через рівні проміжки часу. В ідеальному випадку дискретний сигнал $x(t_n)$ є послідовністю імпульсів нульової тривалості, величина яких дорівнює значенням сигналу в моменти часу nT_s (рис. 1.10, а). Математично такий сигнал можна представити у вигляді:

$$\begin{aligned} x[n] = x(t_n) = x(nT_s) &= \sum_{n=-\infty}^{\infty} x(t) \cdot \delta(t - nT_s) = \\ &= \sum_{n=-\infty}^{\infty} x(nT_s) \cdot \delta(t - nT_s). \end{aligned} \quad (1.6)$$

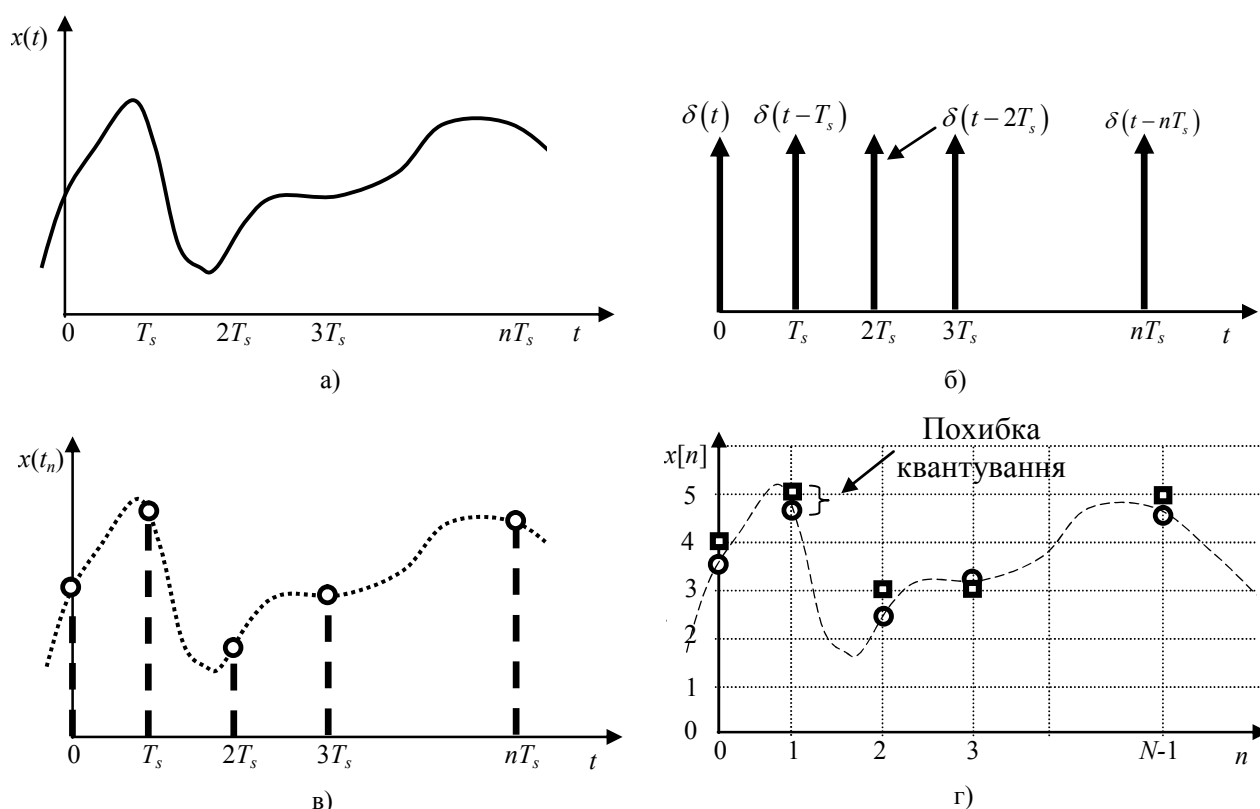


Рисунок 1.10 – Ілюстрація етапів аналого-цифрового перетворення: а) початковий неперервний сигнал; б) послідовність дельта-функцій Дірака в моменти відліків; в) дискретизований сигнал; г) квантований сигнал

В цьому виразі $\delta(t)$ – послідовність функцій Кронекера, які зміщені в ті моменти часу, коли ведеться вимірювання сигналу $x(t)$ (рис. 1.10, б). Функція Кронекера задається так:

$$\delta(t) = \begin{cases} 1, & t = 0, \\ 0, & t \neq 0. \end{cases}$$

2. Квантування.

Аналого-цифрове перетворення потрібне для того, щоб представити значення сигналу, який досліджується, у вигляді, зручному для аналізу в обчислювальній системі. На етапі дискретизації було зроблено перший крок – з неперервного сигналу було залишено лише його значення в певні моменти часу.

Відлік сигналу в загальному випадку є дійсним числом, яке може мати нескінченну кількість знаків після коми. Через це воно не може бути точно представлене в цифровій системі, оскільки цифрова система може працювати лише з числами, що мають скінченну кількість розрядів. Для того, щоб представити довільне число в цифровій системі, роблять наступне.

Дискретизація сигналу у часі при аналого-цифровому перетворенні супроводжується квантуванням за величиною його миттєвих значень у моменти відліків. При цьому весь діапазон значень вхідного сигналу АЦП розбивається на ціле число інтервалів рівної тривалості (квантів, кроків квантування).

Крок квантування – величина інтервалу, на які розбитий динамічний діапазон вхідного сигналу АЦП.

Квантування сигналу – це представлення величини кожного відліку дискретного сигналу у вигляді цілого числа кроків квантування (квантів).

При квантуванні спочатку весь діапазон можливих значень вхідного сигналу АЦП розбивається на цілу кількість дискретних значень (кроків квантування). Від того, яким обирається крок квантування, залежить, наскільки точно ми зможемо представити значення початкового неперервного сигналу. Чим меншим є крок квантування Δ_x (чим на більше число кроків квантування розділений весь діапазон АЦП), тим точнішим є представлення величин дискретних відліків у квантованому сигналі. До процедури квантування значення кожного відліку дискретного сигналу могло бути довільним дійсним числом $x(t_n) \in \mathbb{R}$. При квантуванні в АЦП проводиться заміна цього числа на таке, яке рівне цілій кількості кроків квантування $x(t_n) \rightarrow k\Delta_x$. Можна сказати, що крок квантування – це мінімальне значення «приросту» сигналу, присутність якого в відліку може бути зареєстроване даним АЦП.

3. Кодування.

Як відомо, всі числа у комп'ютерних системах представляються у двійковій системі числення, і вони мають обмежену кількість розрядів. Для того, щоб мати можливість обробляти зареєстровані дискретні та квантовані сигнали, їх також необхідно представити у вигляді двійкового коду. Для цього виконується наступний етап аналого-цифрового перетворення – кодування.

Кодування – представлення *кількості* квантів, якій дорівнює значення сигналу у певний момент часу, за допомогою двійкового коду.

Одною з важливих характеристик АЦП (разом з частотою дискретизації) є *розрядність* або *роздільна здатність* – кількість двійкових розрядів (біт), яка використовується для представлення величини одного відліку. Значення кожного відліку квантується в один з 2^B рівнів. Дискретні значення величини кожного відліку кодуються у вигляді двійкових слів, кожне з яких має довжину B двійкових розрядів.

Наприклад, маючи 5 двійкових розрядів, можна записати 2^5 двійкових числа, від 00000 до 11111 в двійковій системі, або від 0 до 31 в десятковій. Отже, ми можемо за допомогою 5-ти бітного АЦП отримувати цифрові відліки сигналу величиною 0, Δ_x , $2\Delta_x$, $3\Delta_x$, ..., $31\Delta_x$.

Для того, щоб дізнатися значення відліку сигналу в вольтах, треба знати значення кванту. Щоб його визначити, треба знати ще одну характеристику АЦП – діапазон вхідних значень ΔU , В. Якщо АЦП має розрядність B біт, то величина кроку квантування буде дорівнювати $\Delta_x = \frac{\Delta U}{2^B}$, В. Значення відліку сигналу $k\Delta_x$ після квантування буде рівним $x(t_n) = k \frac{\Delta U}{2^B}$.

В разі, коли значення сигналу не дорівнює цілій кількості кроків квантування, з'являється похибка квантування. В результаті квантування отримаємо квантоване значення, яке є найближчим до реального значення дискретного сигналу. В цьому випадку цифровий код на виході АЦП не точно відображає величину значення відліку, а рівний відображає найближчу до нього величину, і яка рівна *цілій* кількості квантів. Похибка буде тим меншою, чим меншим є крок квантування (чим більша розрядність АЦП).

На рисунку 2.4, г реальні значення початкового дискретного сигналу показані кружечками, а отримані квантовані значення – квадратики.

В наступній таблиці показано етапи аналого-цифрового перетворення сигналу з рис. 2.4 для випадку 4-х розрядного АЦП. Тут вважається, що крок квантування дорівнює 0.1 В, а квантування ведеться з надлишком. При цьому округлення ведеться в сторону найближчої більшої цілої кількості квантів.

Початкове значення в вольтах	Номер відліку	Дискретизоване значення в вольтах	Квантоване значення	Кількість квантів, яка підлягає кодуванню	Кодоване значення сигналу
0,35	0	0,35	4 Δ	4	0100
0,48	1	0,48	5 Δ	5	0101
0,24	2	0,24	3 Δ	3	0011
0,61	3	0,61	7 Δ	7	0111

Initial value, Volts	Sample nuber	Discrete value, Volts	Quantized value	Number of quants to be encoded	Code for signal sample
0,35	0	0,35	4 Δ	4	0100
0,48	1	0,48	5 Δ	5	0101
0,24	2	0,24	3 Δ	3	0011
0,61	3	0,61	7 Δ	7	0111

Аналого-цифрове перетворення – найбільш суттєвий етап цифрової обробки сигналів, до якого наразі висувають все більш жорсткі вимоги щодо частоти дискретизації, роздільної здатності та точності перетворення. Зробимо кілька зауважень:

1. В загальному випадку представлення неперервного сигналу набором дискретних значень призводить до *втрати інформації*. Дослідник прагне вивчати реальний, неперервний сигнал, оскільки саме він несе найбільше інформації про об'єкт дослідження. Але якщо до роботи залучаються цифрові системи, то приходится працювати з послідовностями чисел, які представляють початковий реальний сигнал не точно, а є наближенням до нього. Це відбувається тому що при дискретизації втрачається інформація про те, якою була величина сигналу в ті моменти часу, що лежать між відліками. А при квантуванні проводиться заміна реального значення величини відліку на таке, яке дорівнює цілій кількості квантів. Ані похибку дискретизації, ані похибку квантування не можна довільно зменшити без накладання додаткових обмежень на систему цифрової обробки сигналів.

2. Операція аналого-цифрового перетворення сигналу є *незворотною*. В загальному випадку неможливо відновити початковий неперервний сигнал за його цифровими відліками, бо існує нескінченна кількість неперервних сигналів, які матимуть ті значення в моменти відліків, які має наш цифровий сигнал.

Тим не менш, за умови накладання певних обмежень на неперервний сигнал, його теоретично можна точно відновити з дискретного. Про це буде вестись мова в подальших розділах, присвячених частотному представленню дискретизації.

3. Значення дискретного сигналу між відліками *невідоме*. Не можна сказати, що дискретний сигнал там дорівнює нулю, бо він там не існує, оскільки ми самі за допомогою АЦП вимірюємо значення неперервного сигналу лише тільки в моменти відліків. У всі інші моменти часу вимірювання не проводять, отже, якою там була величина сигналу, нам невідомо. Якщо необхідно отримати точні знання про значення сигналу у проміжні моменти часу між моментами дискретизації, то необхідно дискретизувати з більшою частотою. Якщо достатньо неточних відомостей про сигнал між відліками – можна застосувати методи апроксимації.

4. Отримана в результаті аналого-цифрового перетворення сигналу послідовність числових значень відліків *не несе інформації про частоту дискретизації*. Маючи лише числа, наприклад 1, -2, 3, 5, 8..., і знаючи, що це – відліки деякого сигналу, неможливо сказати про те, з яким інтервалом вони були виміряні, або якою була частота дискретизації. Та послідовність чисел, яка для дослідника представляє початковий неперервний сигнал, може бути відліками різних сигналів, які виміряні з різною частотою дискретизації, просто мали однакові значення в моменти відліків. Інформацію про те, як часто проводилися вимірювання величини сигналу (або про крок дискретизації), повинен додатково мати дослідник.

5. Одне з головних питань, яке треба вирішити при проведенні аналого-цифрового перетворення – *як обрати частоту дискретизації*, щоб зберегти максимальний обсяг інформації про початковий неперервний сигнал. Це потрібно для того, щоб за дискретними відліками сигналу можна було робити висновки про властивості початкового неперервного сигналу. В загальному випадку для довільного сигналу це питання не має свого вирішення. Але для певного класу сигналів, а саме для тих, які мають обмежений спектр, за теоремою Котельникова-Шеннона можна визначити

мінімальні вимоги до частоти дискретизації, при якій теоретично можна точно відновити неперервний сигнал з його дискретних відліків.

В подальшому ми будемо в основному працювати лише з неперервними та дискретними сигналами.

1.5. Цифрові послідовності

Математично ми представляємо дискретні сигнали у вигляді послідовностей чисел (рис. 1.11). Послідовність чисел, яка отримана в результаті дискретизації неперервного сигналу $x(t)$, будемо позначати $x[n]$, де n – номер відліку. Якщо розглядається нескінченний сигнал, то номер відліку може приймати значення від мінус нескінченності до нескінченності. Якщо ж маємо скінчений неперервний сигнал, для якого в результаті дискретизації отримано N відліків, то номер відліку $x[n]$ прийматиме значення від 0 до $N - 1$.

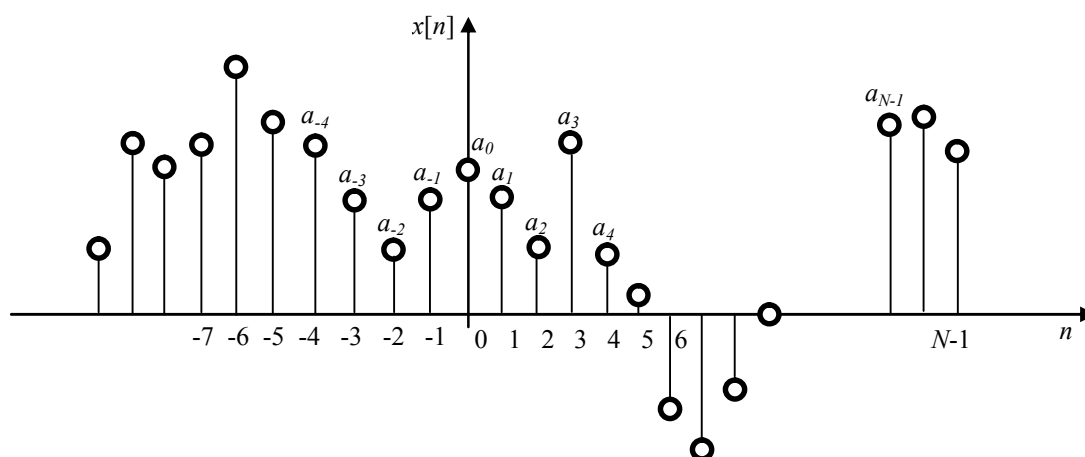


Рисунок 1.11 – Зображення дискретного сигналу

Сигнал $x[n]$ заданий у вигляді чисел – своїми відліками з номерами n . Значення відліку сигналу $x[n]$, з номером 0 дорівнює числу a_0 і записується у вигляді $x[0] = a_0$, значення відліку з номером -2 дорівнює числу a_{-2} : $x[-2] = a_{-2}$, значення відліку з номером 4 дорівнює числу a_4 , а значення відліку з номером $N-1$: $x[N-1] = a_{N-1}$. Часто ми будемо використовувати у математичних виразах поточний відлік сигналу, надаючи йому номер n (n -ий член послідовності $x[n]$). В цьому випадку позначення $x[n]$ буде відноситись і до конкретного відліку, і до всього сигналу в цілому, що не повинно викликати труднощів.

Треба ще раз наголосити, що хоча вісь абсцис на графіку дискретного сигналу (рис. 1.11) неперервна, сигнал заданий лише своїми значеннями для відліків з відповідними номерами. Між відліками сигнал $x[n]$ не існує. Також зазвичай

вважають, що номери сигналів з номерами, меншими за нуль, дорівнюють нулю, тобто $x[n] = 0 \quad \forall n < 0$. Це пішло за аналогією з неперервними сигналами. Вважають, що сигнали починають вимірювати в момент часу $t = 0$. Для неперервного сигналу перше отримане значення буде мати номер $n = 0$.

Над дискретними сигналами (послідовностями чисел) можна виконувати різні математичні операції.

Сума та добуток двох дискретних сигналів визначається почленно: позначення $z[n] = x[n] \cdot y[n]$ говорить про те, що кожний n -ий відлік сигналу $z[n]$ визначається як добуток тих відліків сигналів $x[n]$ та $y[n]$, які також мають номери n .

Наприклад,

$$z[-1] = x[-1] \cdot y[-1], \quad z[0] = x[0] \cdot y[0], \quad z[1] = x[1] \cdot y[1], \quad z[2] = x[2] \cdot y[2], \dots$$

Сума двох сигналів також визначається поелементно: кожний відлік суми дорівнює сумі відліків сигналів-доданків з відповідними номерами:

$$w[n] = x[n] + y[n]:$$

$$w[-1] = x[-1] + y[-1], \quad w[0] = x[0] + y[0], \quad w[1] = x[1] + y[1], \dots$$

Добутком сигналу на число є сигнал, в якому всі відліки отримуються в результаті добутку кожного відліку на це число.

Сигнал $y[n]$ називається затриманою або зсунутою версією сигналу $x[n]$, якщо виконується рівність

$$y[n] = x[n - n_0],$$

де n_0 – ціле число.

Наприклад:

Нехай маємо сигнал, що складається з п'яти відліків $x[n] = [2, -1, 3, 8, 9]$. Затриманий на 2 відліки сигнал $y[n] = x[n - 2]$ буде визначатися так.

$y[0] = x[0 - 2] = x[-2] = 0$ (нагадаємо, що відліки сигналу нумеруються починаючи з нуля, а всі відліки з нульовими номерами вважаються нульовими).

$$y[1] = x[1 - 2] = x[-1] = 0$$

$$y[2] = x[2 - 2] = x[0] = 2$$

$$y[3] = x[3 - 2] = x[1] = -1$$

$$y[4] = x[4 - 2] = x[2] = 3$$

$$y[5] = x[5 - 2] = x[3] = 8$$

$$y[6] = x[6 - 2] = x[4] = 9$$

$$y[7] = x[7 - 2] = x[5] = 0$$

$$y[8] = x[8 - 2] = x[6] = 0$$

і т.д.

Дискретний сигнал $x[n]$ є періодичним, якщо існує таке ціле число N , яке називається періодом, що виконується рівність:

$$x[n \pm kN] = x[n], \text{ де } k - \text{ціле число.}$$

Розглянемо деякі важливі послідовності.

Для дискретних сигналів аналогом дельта-функції Дірака є *одиничний імпульс* (рис. 1.12 – 1.13):

$$\delta[n] = \begin{cases} 0, & n \neq 0, \\ 1, & n = 0. \end{cases} \quad (1.7)$$

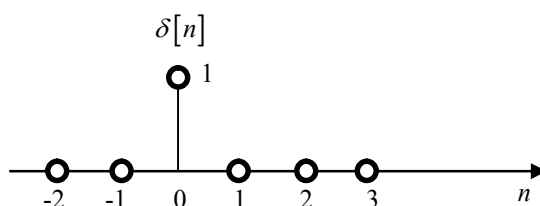


Рисунок 1.12 – Одиничний імпульс

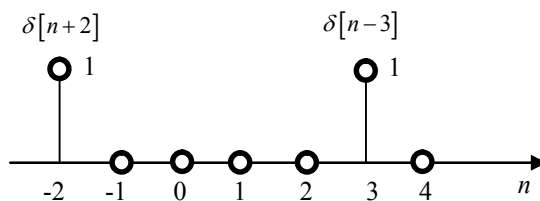


Рисунок 1.13 – Затримані одиничні імпульси

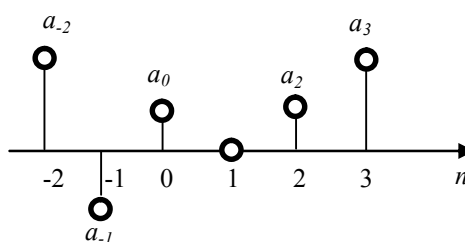


Рисунок 1.14

Корисним є те, що з використанням одиничного імпульсу можна виразити будь-який дискретний сигнал $x[n]$ у вигляді лінійної комбінації зсунутих одиничних імпульсів. Наприклад, сигнал $x[n]$ з шести відліків (рис. 1.14) можна представити у вигляді:

$$x[n] = a_{-2}\delta[n+2] + a_{-1}\delta[n+1] + a_0\delta[n] + a_2\delta[n-2] + a_3\delta[n-3].$$

Для довільного дискретного сигналу можна записати:

$$x[n] = \sum_{k=-\infty}^{+\infty} x[k] \delta[n-k]. \quad (1.8)$$

Цей вираз є аналогом фільтруючої властивості неперервної дельта-функції Дірака. Наприклад, якщо маємо сигнал $x[n] = [2, -1, 3, 8, 9]$, то його запис через зсунуті одиничні імпульси матиме такий вигляд:

$$\begin{aligned} x[n] &= \sum_{k=0}^4 x[k] \delta[n-k] = x[0] \delta[n] + x[1] \delta[n-1] + \\ &+ x[2] \delta[n-2] + x[3] \delta[n-3] + x[4] \delta[n-4] = \\ &= 2\delta[n] - \delta[n-1] + 3\delta[n-2] + 8\delta[n-3] + 9\delta[n-4]. \end{aligned}$$

Маючи цей вираз, можна розрахувати значення будь-якого відліку сигналу $x[n]$: для цього треба в вираз підставити значення номеру n провести розрахунок (враховуючи значення одиничної послідовності з (1.7)). Наприклад,

$$\begin{aligned} x[0] &= 2\delta[0] - \delta[0-1] + 3\delta[0-2] + 8\delta[0-3] + 9\delta[0-4] = \\ &= 2\delta[0] - \delta[-1] + 3\delta[-2] + 8\delta[-3] + 9\delta[-4] = \\ &= 2 \cdot 1 - 1 \cdot 0 + 3 \cdot 0 + 8 \cdot 0 + 9 \cdot 0 = 2 \cdot 1 = 2. \end{aligned}$$

$$\begin{aligned} x[3] &= 2\delta[3] - \delta[3-1] + 3\delta[3-2] + 8\delta[3-3] + 9\delta[3-4] = \\ &= 2\delta[3] - \delta[2] + 3\delta[1] + 8\delta[0] + 9\delta[-1] = \\ &= 2 \cdot 0 - 1 \cdot 0 + 3 \cdot 0 + 8 \cdot 1 + 9 \cdot 0 = 8 \cdot 1 = 8. \end{aligned}$$

$$\begin{aligned} x[15] &= 2\delta[15] - \delta[15-1] + 3\delta[15-2] + 8\delta[15-3] + 9\delta[15-4] = \\ &= 2\delta[15] - \delta[14] + 3\delta[13] + 8\delta[12] + 9\delta[11] = \\ &= 2 \cdot 0 - 1 \cdot 0 + 3 \cdot 0 + 8 \cdot 0 + 9 \cdot 0 = 0. \end{aligned}$$

Одинична послідовність (рис. 1.15) визначається формулою

$$u[n] = \begin{cases} 1, & n \geq 0, \\ 0, & n < 0. \end{cases}$$

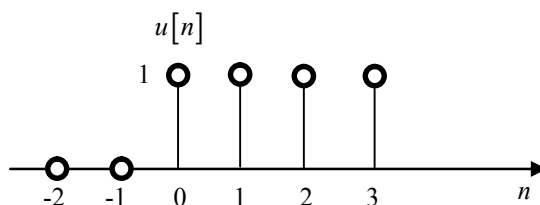


Рисунок 1.15 – Одинична послідовність

Одиничну послідовність (як і будь-яку іншу) також можна виразити через одиничний імпульс:

$$u[n] = \sum_{k=0}^{\infty} \delta[n-k].$$

Ця послідовність є аналогом функції Хевісайда для дискретних сигналів.

Експоненціальна послідовність

$$x[n] = A\alpha^n.$$

Якщо A та α є дійсними числами, то послідовність також дійсна. Якщо $0 < \alpha < 1$ і A додатне, то значення послідовності додатні і зменшуються з ростом n . Якщо $-1 < \alpha < 0$, знаки членів послідовності чергуються, а абсолютні значення зменшуються. Якщо $|\alpha| > 1$, то послідовність зростає за абсолютною величиною.

Синусоїдальна послідовність

$$x[n] = A\cos(\omega_0 n + \varphi).$$

Є дискретним аналогом неперервного синусоїдального сигналу. Дискретні синусоїдальні сигнали відіграють важливу роль при розкладі сигналів за Фур'є, тому такі сигнали більш докладно будуть розглядатися у подальшому. Цікавим є факт, що дискретний синусоїдальний сигнал $x[n] = A\sin(n)$ є неперіодичною функцією, на відміну від його неперервного аналога $x(t) = A\sin(t)$. Це пов'язано з тим, що період дискретного сигналу має бути цілим числом (оскільки значення дискретного сигналу задані лише для цілих значень аргументу). А число 2π не є цілим.

1.6. Простори сигналів

Сигнал, який залежить від однієї змінної, наприклад, від часу, у графічному представленні зображується сукупністю точок, які разом складають криву у двовірному просторі (рис. 1.1). Існує багато математичних методів дослідження таких функцій. Математичний апарат лінійної алгебри і аналітичної геометрії дозволяє узагальнити описання сигналів з різними властивостями. Надалі ми розглянемо більш складні способи представлення сигналів, при яких кожний сигнал (неперервна функція чи послідовність) зображується лише одною точкою в деякому просторі. Залежно від того, якими є властивості сигналів, треба обирати той математичний простір, елементи якого мають такі самі властивості.

Для того, щоб сформулювати загальний підхід до аналізу сигналів, в теорії сигналів вводиться поняття «простір сигналів».

Простір – це множина елементів будь-якої природи, яка має структуру, тобто для елементів якої визначені деякі операції, або між елементами якої встановлені деякі співвідношення. Зазвичай визначають операції додавання або віднімання сигналів, та масштабування сигналів (множення на скаляр).

Простір сигналів – множина, елементами якої є неперервні або дискретні сигнали. Далі будуть наведені основні відомості з лінійної алгебри, які використовуються у теорії сигналів.

1.6.1. Лінійний простір сигналів

Для того, щоб множина сигналів була простором, вона повинна бути замкненою: тобто зважена сума елементів простору також має бути елементом цього простору.

Лінійний (афінний) простір сигналів – множина сигналів x, y, \dots , які наділені векторними властивостями, тобто для них визначені операції додавання та множення на скаляр. Елементи такого простору називаються векторами, вони можуть бути будь-якими об'єктами, зокрема, сигналами (рис. 1.16). Лінійні простори можуть бути скінченними, зліченими або нескінченними.

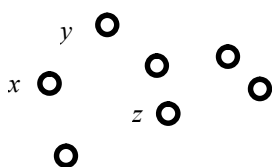


Рисунок 1.16 – Вектори у просторі

Для такої множини виконуються три вимоги:

1. Існує правило, за яким двом елементам цього простору ставиться у відповідність третій елемент цього ж простору, який називається сумою елементів:

$$x + y = z.$$

2. Існує правило, за яким будь-якому елементу цього простору та будь-якому дійсному числу λ ставиться у відповідність елемент цього ж простору, який називається добутком елемента x на число λ та позначається λx .

3. Вказані два правила підкоряються таким аксіомам:

- $x + y = y + x$ (векторна комутативність);
- $x + (y + z) = (x + y) + z$ (векторна асоціативність);
- в просторі існує єдиний нульовий елемент, такий що $x + 0 = x$;
- унітарність множення на скаляр: $1 \cdot x = x$;
- для кожного елемента x існує єдиний протилежний елемент $-x$, такий що $x + -x = 0$;
- $\alpha \cdot (\beta x) = (\alpha \beta) x$ для будь-яких скалярів α, β та елемента x (скалярна асоціативність);
- $\alpha(x + y) = \alpha x + \alpha y$ для будь-якого скаляра α (векторна дистрибутивність);
- $(\alpha + \beta)x = \alpha x + \beta x$ (скалярна дистрибутивність);

Зазвичай як скаляри використовуються дійсні або комплексні числа.

Для неперервних сигналів операції суми та добутку на скаляр визначаються природнім чином: додавання є поточечним додаванням значень сигналу у співпадаючі моменти часу, а множення на скаляр зводиться до множення функції, що описує сигнал, на скалярну константу.

Лінійною комбінацією векторів дійсного лінійного простору $\{x_1, x_2, \dots, x_n\}$ називається *вектор*, утворений як сума цих векторів з деякими скалярними коефіцієнтами:

$$y = \sum_{i=1}^n a_i x_i. \quad (1.9)$$

Для кожного набору з n скалярних коефіцієнтів $\{a_i\}$ буде отриманий інший результуючий вектор y (лінійна комбінація). Множина всіх можливих лінійних комбінацій векторів з лінійного простору L (тобто з множини $\{x_1, x_2, \dots, x_n\}$) які можуть бути отримані за (1.9), також утворює лінійний простір.

Лінійною оболонкою елементів лінійного простору $\{x_1, x_2, \dots, x_n\}$ називається сукупність всіх лінійних комбінацій цих елементів, тобто всі можливі вектори y , які утворені за (1.9) з довільними числами $\{a_i\}$.

Підмножина L елементів лінійного простору R називається **лінійним підпростором** простору R , якщо вона задовольняє умовам:

- якщо елементи x та y належать до підмножини L , то і їх сума також належить підмножині L ;
- якщо елемент x належить підмножині L , і λ – деяке дійсне число, то елемент λx також належить підмножині L .

Якщо взяти підмножину $\{x_1, x_2, \dots, x_m\}$, де $m < n$, елементів лінійного простору, то всі лінійні комбінації елементів цієї підмножини також утворюють лінійний простір, який є підпростором початкового лінійного простору, що утворений лінійними комбінаціями початкових елементів $\{x_1, x_2, \dots, x_n\}$.

Лінійна оболонка є найменшим підпростором, який містить всі елементи $\{x_1, x_2, \dots, x_n\}$, з яких вона утворена. Говорять, що такий підпростір-лінійна оболонка є «натягнутою» на елементи $\{x_1, x_2, \dots, x_n\}$.

Елементи $\{x_1, x_2, \dots, x_n\}$ називаються **лінійно незалежними**, якщо лінійна комбінація цих елементів з дійсними числами $\{a_i\}$ дорівнює нульовому елементу $\sum_{i=1}^n a_i x_i = 0$

тільки при умові $a_1 = a_2 = \dots = a_n = 0$, $x_n \neq 0$. Якщо можна знайти такі дійсні числа $\{a_i\}$, з яких хоча б одне *не* дорівнює нулю, що лінійна комбінація дорівнює нульовому елементу – такі елементи називаються **лінійно залежними**. Іншими словами, в лінійно-незалежній множині жоден елемент не може бути представлений

у вигляді лінійної комбінації інших елементів цієї ж самої множини (не може бути сумою інших елементів цієї ж множини). І навпаки, якщо ми маємо лінійно-залежну множину, то в ній один вектор може бути виражений у вигляді лінійної комбінації інших:

$$x_k = \sum_{\substack{i=1 \\ i \neq k}}^N a_i x_i.$$

Набір векторів $\{x_1, x_2, \dots, x_n\}$ називається повним, якщо будь-який вектор y , що належить даному простору, можна представити у вигляді лінійної комбінації векторів цього простору.

Якщо розглядаємо нескінченновимірну систему векторів, то вона буде лінійно-незалежною у випадку, якщо лінійно-незалежною буде будь-яка її скінченновимірна підсистема.

Нехай простір M – лінійна оболонка для N лінійно-незалежних векторів $\{x_i\} = \{x_1, x_2, \dots, x_N\}$. Кожний вектор y в просторі M може бути представлений у вигляді лінійної комбінації векторів $\{x_i\}$ єдиним способом, тобто кожному вектору з цієї множини відповідає лише один набір скалярних коефіцієнтів $\{a_i\}$:

$$y = \sum_{i=1}^N a_i x_i.$$

Лінійний простір називається **N -вимірним**, якщо в ньому існує N лінійно незалежних елементів, а будь-які $N+1$ елементів вже є лінійно-залежними. При цьому число N називають **розмірністю** простору.

Сукупність *будь-яких* N лінійно-незалежних елементів $\{e_i\}$ N -вимірного простору L називається **базисом** цього простору. Базис має такі властивості:

1. Будь-який вектор з простору L може бути виражений як лінійна комбінація базисних векторів. Рівність

$$y = \sum_{i=1}^N a_i e_i = a_1 e_1 + a_2 e_2 + \dots + a_N e_N \quad (1.10)$$

називається **розкладом** довільного елемента y простору L в базисі $\{e_1, e_2, \dots, e_N\}$, а числа $\{a_1, a_2, \dots, a_N\}$ – координатами елемента y відносно базису $\{e_1, e_2, \dots, e_N\}$.

2. Лінійною оболонкою базисних векторів деякого простору є сам цей простір.

Вираз (1.10) є одним з найбільш важливих в теорії сигналів. Він показує в загальному вигляді, що можна представити наш довільний сигнал y в деякому просторі за допомогою сукупності базисних елементів цього простору $\{e_1, e_2, \dots, e_N\}$.

Ці елементи – також сигнали, але спеціально підібрані таким чином, щоб представляти з їх допомогою довільні сигнали було зручно та/або корисно. Як

правило, базисні сигнали – це деякі прості сигнали з відомими та характерними властивостями.

Аналізуючи представлення сигналу (1.10), ми можемо дослідити властивості нашого сигналу через властивості тих базисних елементів $\{e_1, e_2, \dots, e_N\}$, що його складають. Тим, наскільки великим буде значення коефіцієнту a_i в розкладі сигналу, буде визначатися, наскільки значним буде вклад складової e_i в досліджуваний сигнал y . По величині коефіцієнтів розкладу $\{a_1, a_2, \dots, a_n\}$ можна здобути відомості щодо складу сигналу (1.10). Якщо деякі з коефіцієнтів більші за інші, то це буде означати, що в сигналі присутні характерні властивості, які відображені у відповідних складових розкладу. А отже, знаючи властивості складових $\{e_1, e_2, \dots, e_N\}$ та їх вираженість в представленні сигналу (величини коефіцієнтів $\{a_1, a_2, \dots, a_N\}$), можна зробити висновки про властивості самого сигналу. Можна сказати, що замість того, щоб працювати зі складними об'єктами – векторами, ми можемо працювати з більш простими об'єктами – скалярами, які є коефіцієнтами розкладу сигналу в базисі. Підбираючи базиси на основі того, які саме властивості ми хочемо знайти (виявити) в сигналі, та розраховуючи коефіцієнти розкладу, ми можемо проводити аналіз та вивчення сигналів, а отже, і тих об'єктів, від яких сигнал надійшов. Методи вибору базисів розкладу з існуючих, а також методи побудови базисів, використання яких при розкладах буде виявляти в сигналах певні цікаві досліднику властивості – окремий розділ теорії сигналів.

Розглянемо ілюстрацію просторів з заданими базисами. На рис. 1.17 приклад двовимірного простору. В ньому задані два базисних елементи $\{e_1, e_2\}$ і всі інші вектори цього простору можуть бути виражені у вигляді розкладу на ці два базисних вектора:

$$\begin{aligned}x &= ae_1 + be_2, \\y &= ce_1 + de_2.\end{aligned}$$

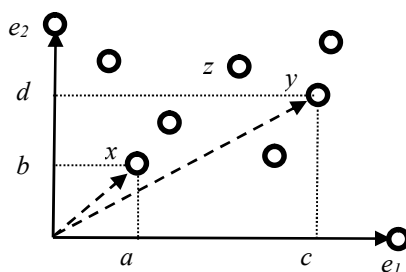


Рисунок 1.17 – Двовимірний простір сигналів

Вектори x та y в цьому базисі мають координати $x = [a, b]$ та $y = [c, d]$. Ці вектори зображені кружечками та додатково з метою кращої наочності – направленими відрізками, що виходять з початку координат. Початок координат – це точка в цьому просторі з координатами $[0, 0]$ відносно обраного базису $\{e_1, e_2\}$.

Якщо простір тривимірний, то в кожного вектора будуть три координати відносно базисних векторів $\{e_1, e_2, e_3\}$, наприклад (рис. 1.18):

$$x = ae_1 + be_2 + ce_3.$$

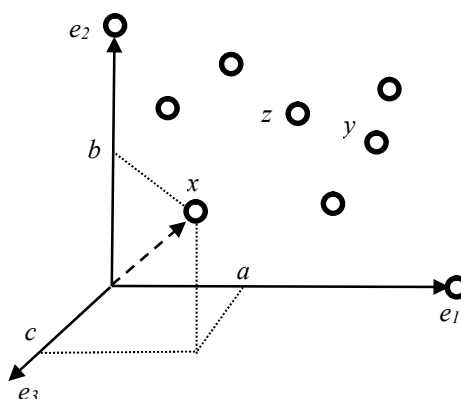


Рисунок 1.18 – Тривимірний простір сигналів

Якщо сигнал належить простору з більшою розмірністю, то він має більше координат.

Розглянемо приклади просторів, елементами яких є найпоширеніші сигнали.

ПРИКЛАД 1

Множина всіх впорядкованих сукупностей з N дійсних або комплексних чисел (N -вимірних векторів-рядків) в просторі утворює N -вимірний лінійний векторний простір.

Нехай $x = \{a_1, a_2, \dots, a_N\}$ а $y = \{b_1, b_2, \dots, b_N\}$. Тоді додавання векторів можна визначити у вигляді $x + y = \{a_1 + b_1, a_2 + b_2, \dots, a_N + b_N\}$, а добуток – $x \cdot y = \{a_1 b_1, a_2 b_2, \dots, a_N b_N\}$.

Ясно, що будь-який елемент цього простору може бути представлений лінійною комбінацією

$$x = \sum_{i=1}^N e_i x_i = e_1 x_1 + e_2 x_2 + \dots + e_N x_N,$$

де N лінійно незалежних векторів задаються так:

$$e_1 = \underbrace{\{1, 0, 0, \dots, 0\}}_N$$

$$e_2 = \underbrace{\{0, 1, 0, \dots, 0\}}_N$$

...

$$e_N = \underbrace{\{0, 0, 0, \dots, 1\}}_N$$

Елементами в такому просторі є дискретні сигнали, кожний з яких складається з N відліків. Множина векторів $\{e_i\}_{i=1}^N$ називається евклідовим базисом. В цьому просторі існують також багато інших базисів, про які буде іти мова далі.

ПРИКЛАД 2

Множина всіх дійсних або комплексних функцій часу, визначених на інтервалі $T = \{t, a \leq t \leq b\}$, утворює лінійний простір, елементами якого є неперервні сигнали, обмежені за тривалістю. Додавання векторів можна визначити у вигляді $z(t) = x(t) + y(t)$, а добуток – поточечним множенням значень функцій на всьому інтервалі.

1.6.2. Метричний простір сигналів

Часто при дослідженні не одного, а кількох сигналів виникає задача порівняти пару сигналів та визначити, наскільки вони подібні чи схожі один на одного. Це може бути ознакою того, що ці сигнали спричинені схожими процесами в об'єкті, або процесами, що мають одну причину. Математичним еквівалентом такого порівняння сигналів є визначення відстаней між тими функціями, що описують сигнали. Якщо відстань невелика, то можна вважати, що сигнали «близькі», а отже – схожі. Якщо ж значення відстані між функціями достатньо велике, то можна зробити висновок, що і самі сигнали не схожі один на одного. Отже, необхідно мати математичний метод для співставлення у відповідність парі сигналів деякого числа, яке буде характеризувати їх близькість – відстань між ними.

Для того, щоб задати відстань між сигналами, не достатньо працювати тільки в лінійному просторі, оскільки в ньому задані тільки операції додавання двох сигналів та множення сигналу на скаляр. Для розрахунку схожості сигналів треба доповнити лінійний простір сигналів потрібною нам операцією – перейти до метричного простору сигналів.

Метричний простір сигналів – це простір сигналів, між якими можна розраховувати відстань за допомогою метрики. Метричний простір утворюється лінійним простором сигналів разом з метрикою.

Метрика – це правило, яке задає відображення впорядкованої пари сигналів (x, y) у дійсне число $D(x, y)$. Метрику для двох сигналів ще називають відстанню між ними. Відображення елементів простору у число називається функціоналом, тож метрика – це функціонал.

Для множини сигналів може бути багато способів визначення відстані між сигналами. Відстань повинна мати такі властивості:

1. Невід'ємність відстані: $D(x, y) \geq 0$, причому $D(x, y) = 0$ тільки при $x = y$.
2. Симетрія відстані: $D(x, y) = D(y, x)$.
3. Нерівність трикутника: $D(x, z) \leq D(x, y) + D(y, z)$.
4. Умова ідентичності (тотожності): $D(x, y) = 0 \Leftrightarrow x = y$

Якщо умова ідентичності не виконується при виконанні трьох інших, то функція $D(x, y)$ називається *псевдометрикою*. Нерівність трикутника формалізує вимогу, що якщо дві точки простору близькі до третьої, то вони повинні бути близькими і одна до одної.

На одному і тому самому просторі можна задати багато метрик. Для векторів N -вимірному векторного простору дійсних або комплексних дискретних сигналів можна ввести p -метрику:

$$d_p(x, y) = \sqrt[p]{\sum_{n=0}^{N-1} |x[n] - y[n]|^p}. \quad (1.11)$$

Зокрема, при $p = 2$ маємо евклідову метрику:

$$d_2(x, y) = \sqrt{\sum_{n=0}^{N-1} |x[n] - y[n]|^2}.$$

Можна ввести метрику Хеммінга (манхетенську метрику, або відстань Мінковського):

$$d_1(x, y) = \sum_{n=0}^{N-1} |x[n] - y[n]|.$$

Також можна використати метрику рівномірного наближення (метрику Чебишова):

$$d_\infty(x, y) = \max_{n=0 \dots N-1} |x[n] - y[n]|.$$

Сказане вище можна розповсюдити на нескінченновимірний простір дискретних сигналів. Але наведені метрики можна використовувати лише у випадку, коли виконується умова збіжності для сигналів:

$$\sum_{n=0}^{\infty} x^2[n] < \infty$$

Для сигналів, які описуються неперервними функціями часу на інтервалі T , можна також ввести p -метрику:

$$d_p(x, y) = \sqrt[p]{\int_T |x(t) - y(t)|^p dt}. \quad (1.12)$$

Аналогічно до випадку аналізу дискретних сигналів, частіше за все використовують значення $p=1, 2$ та ∞ .

При цьому загальна формула приймає вигляд:

$$\begin{aligned} d_1(x, y) &= \int_T |x(t) - y(t)| dt, \\ d_2(x, y) &= \sqrt{\int_T |x(t) - y(t)|^2 dt}, \\ d_\infty(x, y) &= \max_{t \in T} |x(t) - y(t)|. \end{aligned}$$

При $p = 2$ метрика називається евклідовою. Вона використовується в теорії сигналів найчастіше, оскільки зручна для розрахунку та її величині можна надати певний фізичний зміст: її величина пропорційна енергії різниці двох сигналів. Метричний простір з евклідовою метрикою будемо називати *евклідовим простором*.

При виборі метрики в інтегральному вигляді, відстань між двома сигналами рівна нулю, якщо ці сигнали є рівними «майже усюди», тобто вони відрізняються один від одного у зліченній множині ізольованих точок (на множині міри нуль). Ця різниця не вносить вклад до інтегралу, за яким розраховують метрику: він має зміст площі фігури, яка утворена при відніманні двох сигналів.

1.6.3. Нормований простір

Наступним удосконаленням структури та можливостей представлення сигналів як елементів просторів є введення простих алгебраїчних взаємозв'язків між сигналами. Часто виникає задача оцінити сигнал, який аналізується, сам по собі. Тобто, однозначно пов'язати з кожним сигналом число. Порівнюючи ці числа між собою, можна дізнатися, як співвідносяться між собою і самі сигнали. Для цього в просторі сигналів додатково вводять *норму* для кожного вектора, отримуючи нормований лінійний простір.

Нормований простір – це простір, в якому додатково визначений «розмір» кожного елементу x . Цей параметр називається нормою або довжиною вектора та позначається $\|x\|$.

Норма вектора – це дійсне число, яке ставиться у відповідність будь-якому елементу простору за деяким правилом. Для цього правила мають виконуватися такі вимоги:

1. Норма повинна бути невід'ємною: $\|x\| \geq 0$, $\|x\| = 0$ тільки при $x = 0$;
2. Нерівність трикутника $\|x + y\| \leq \|x\| + \|y\|$;

3. Для довільного скаляра a виконується рівність $\|a \cdot x\| = |a| \cdot \|x\|$.

Можна побачити взаємозв'язок між нормою та метрикою. Спосіб введення норми може бути узгоджений із введеною в даному просторі метрикою, а саме: метрика повинна представляти собою норму різниці елементів простору $\|x - y\|$. І навпаки, якщо в лінійному просторі задана метрика, то нормою може служити відстань між вектором та нульовим вектором (початком координат) тобто $D(x, 0)$. Отже, наведена раніше p -метрика породжує p -норму:

$$\|x\|_p = \sqrt[p]{\int_T |x(t)|^p dt}. \quad (1.13)$$

Можна побудувати норми, аналогічні побудованим раніше метрикам. Для випадку неперервних сигналів:

$$\begin{aligned} \|x\|_1 &= \int_T |x(t)| dt, \\ \|x\|_2 &= \sqrt{\int_T |x(t)|^2 dt}, \\ \|x\|_\infty &= \max_{t \in T} |x(t)|. \end{aligned}$$

Для випадку дискретних сигналів:

$$\begin{aligned} \|x\|_1 &= \sum_{n=1}^{\infty} |x[n]|, \\ \|x\|_2 &= \sqrt{\sum_{n=1}^{\infty} |x[n]|^2}, \\ \|x\|_\infty &= \max_{n=1 \dots \infty} |x[n]|. \end{aligned}$$

Нормований простір, який є також метричним простором, називають *банаховим простором* (по імені відомого польського математика Стефана Банаха).

В аналізі сигналів часто виникає задача оцінки міри схожості чи відмінності двох сигналів. Для цього можуть використовуватися норми різниць тих векторів, що відповідають даним сигналам в банаховому просторі. Здебільшого використовують варіації степенової норми

$$s = \|x - y\|_{p/r} = \sqrt[r]{\int_T |x(t) - y(t)|^p dt}, \quad (1.14)$$

де p – параметр, який контролює відносну вагу в результуючій величині схожості s тієї розбіжності між значеннями сигналу, яка присутня в кожний момент часу; r – параметр, що контролює величину збільшення відстані (норми різниці) між сигналами при збільшенні реальної віддаленості між сигналами у просторі.

У випадку коли $r = p$ норма (1.14) називається нормою Мінковського, коли $r = p = 1$ – отримується манхеттенська норма, при $r = p = 2$ – евклідова норма, $r = p \rightarrow \infty$ – норма Чебишова.

1.6.4. Гільбертів простір (простір зі скалярним добутком)

Існує дуже поширена група задач, в ході розв'язання яких необхідно мати одну числову характеристику для двох сигналів. Ця необхідність виникає, наприклад, при порівнянні двох сигналів, або при розрахунку проекцій досліджуваного сигналу на базисні вектори. В такому випадку простір, елементами якого є сигнали, треба доповнити ще одною операцією, що є ключовою для всієї обробки сигналів – скалярним добутком. Слід зауважити, що для сигналів як елементів простору поняття скалярного добутку відрізняється від аналогічного поняття, яке вводилось для векторів на площині в геометрії (що є окремим випадком). В теорії сигналів поняття скалярного добутку вводять аксіоматично.

Скалярний добуток $\langle x, y \rangle$ – це спосіб відображення впорядкованих пар векторів простору на комплексну площину, тобто співставлення їм у відповідність одного числа.

Іноді скалярний добуток називають внутрішнім добутком (inner product, dot product, scalar product). Правило, за яким отримують це відображення в просторі M , повинне задовольняти таким властивостям:

1. Невід'ємність $\langle x, x \rangle \geq 0$, $\forall x \in M$, причому $\langle x, x \rangle = 0$ тільки при $x = 0$;
2. Ермітова симетричність $\langle x, y \rangle = \langle y, x \rangle^*$ (тут і далі зірочка відповідає комплексному спряженню). Отже, скалярний добуток є комутативним лише у випадку, коли вектори є дійсними;
3. $\langle ax, y \rangle = a \langle x, y \rangle$ та $\langle x, ay \rangle = a^* \langle x, y \rangle$
4. Дистрибутивність $\langle ax + by, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$;

На основі скалярного добутку можна ввести норму для сигналів в просторі:

$$\|x\|_p = \sqrt[p]{\langle x, x \rangle}.$$

Норма, в свою чергу, породжує метрику:

$$D_p(x, y) = \|x - y\|_p = \sqrt[p]{\langle x - y, x - y \rangle}.$$

Для неперервних та дискретних сигналів скалярний добуток може бути введений за такими правилами

$$\langle x, y \rangle = \int_T x(t) y^*(t) dt, \quad (1.15)$$

$$\langle x, y \rangle = \sum_{n=0}^{N-1} x[n] y^*[n]. \quad (1.16)$$

Скалярні добутки (1.15)-(1.16) породжують евклідові норму та метрику.

Доведемо нерівність Коші-Буняковського (нерівність Шварца) для такого аксіоматичного способу задавання скалярного добутку. Розглянемо скалярний добуток $\langle x + ay, x + ay \rangle$, який згідно з першою властивістю скалярного добутку є невід'ємним: $\langle x + ay, x + ay \rangle \geq 0$. Розкриємо скалярний добуток:

$$\begin{aligned} \langle x + ay, x + ay \rangle &= \\ &= \langle x, x \rangle + \langle x, ay \rangle + \langle ay, x \rangle + \langle ay, ay \rangle \geq 0 \\ \langle x, x \rangle + a^* \langle x, y \rangle + a \langle x, y \rangle^* + |a|^2 \langle y, y \rangle &\geq 0 \end{aligned}$$

Остання нерівність виконується для всіх значень скаляру a . Оберемо значення

$$a = -\frac{\langle x, y \rangle}{\langle y, y \rangle}$$

та підставимо його у вираз:

$$\langle x, x \rangle - \frac{\langle x, y \rangle^*}{\langle y, y \rangle} \langle x, y \rangle - \frac{\langle x, y \rangle}{\langle y, y \rangle} \langle x, y \rangle^* + -\frac{\langle x, y \rangle \langle x, y \rangle^*}{\langle y, y \rangle^2} \langle y, y \rangle \geq 0,$$

і після скорочення будемо мати:

$$\langle x, x \rangle - \frac{\langle x, y \rangle^*}{\langle y, y \rangle} \langle x, y \rangle \geq 0.$$

Помножимо ліву і праву частину нерівності на $\langle y, y \rangle$, причому знак нерівності не поміняється, оскільки це додатне число:

$$\begin{aligned} \langle x, x \rangle \langle y, y \rangle - \langle x, y \rangle^* \langle x, y \rangle &\geq 0 \\ \langle x, x \rangle \langle y, y \rangle - |\langle x, y \rangle|^2 &\geq 0 \Rightarrow \\ |\langle x, y \rangle|^2 &\leq \|x\|^2 \cdot \|y\|^2. \end{aligned}$$

Отже, скалярний добуток двох векторів не перевищує добуток довжин цих векторів. Нерівність Коші-Буняковського використовується, зокрема, при розв'язанні задач оптимізації передавальних функцій лінійних систем.

Розглянемо корисний випадок, коли вектори є *ортогональними*. Вектори називаються ортогональними, якщо їх скалярний добуток дорівнює нулю:

$$\langle x, y \rangle = 0 \Rightarrow x \perp y.$$

Розглянемо деякі властивості ортогональних векторів. Нехай M_N – довільний N -вимірний простір сигналів з ортогональним базисом $\{\varphi_i\}_{i=1}^N$. Визначимо, чому

$$\begin{aligned} \text{дорівнює величина } \left\| \sum_{i=1}^N \varphi_i \right\|^2 : \\ \left\| \sum_{i=1}^N \varphi_i \right\|^2 = \left\langle \sum_{i=1}^N \varphi_i, \sum_{i=1}^N \varphi_i \right\rangle = \langle \varphi_1 + \varphi_2 + \dots, \varphi_1 + \varphi_2 + \dots \rangle = \\ = \langle \varphi_1, \varphi_1 \rangle + \langle \varphi_2, \varphi_2 \rangle + \dots + \langle \varphi_1, \varphi_2 \rangle + \langle \varphi_1, \varphi_3 \rangle + \dots \end{aligned}$$

В останній сумі всі доданки, в яких присутній скалярний добуток різних векторів з ортогонального базису, будуть дорівнювати нулю, отже

$$\left\| \sum_{i=1}^N \varphi_i \right\|^2 = \sum_{i=1}^N \|\varphi_i\|^2.$$

Видно, що квадрат норми суми базисних векторів дорівнює сумі квадратів норм кожного з векторів базису.

Також можна довести, що ортогональна система векторів є лінійно-незалежною.

1.7. Представлення сигналів в гільбертовому просторі

Як правило, сигнали є складними функціями часу або простору. Одним із шляхів аналізу сигналів є представлення у вигляді сукупності деяких більш простих елементарних складових, які або більш придатні для аналізу, або яким можна надати реальний зміст. Під терміном «представлення» (representation) мають на увазі спосіб описання сигналів з допомогою інших функцій або наборів чисел.

Раніше було введено поняття розкладу сигналу в базисі лінійного простору. Зараз ми розвинемо та уточнимо поняття розкладу.

Розглянемо гільбертів простір L^2 і сигнал $x \in L^2$. Нехай M_N – довільний N -вимірний підпростір сигналів з довільним базисом $\{\varphi_i\}_{i=1}^N$. Тобто, простір M_N є натягнутою на цей базис лінійною оболонкою базисних векторів, а вектори базису належать цьому підпростору: $\{\varphi_i\}_{i=1}^N \in M_N$.

Нехай наш сигнал x належить також підпростору M_N : $x \in M_N$.

Відомо, що будь-який сигнал підпростору має єдиний розклад в базисі цього підпростору, тобто може бути представлений у вигляді лінійної комбінації базисних векторів:

$$x = \sum_{i=1}^N c_i \varphi_i, \quad (1.17)$$

де c_i – коефіцієнти розкладу.

Вираз (1.17) показує, що ми можемо *представити* елемент x у базисі у вигляді його розкладу за елементами цього базису. Оскільки представлення (1.17) є однозначним, то можна розглядати замість самого вектора x його розклад, і всі властивості x мають своє відображення у цьому розкладі. Іншими словами, замість дослідження самого сигналу можна проаналізувати коефіцієнти розкладу (координати сигналу в базисі), властивості елементарних сигналів, на які ведеться розклад (базисних векторів). В результаті цього можна отримати відомості про сам сигнал. Це одна з фундаментальних ідей всієї теорії сигналів та методів перетворення сигналів.

Головне питання – як знайти той набір чисел $\{c_i\}$, які дають інформацію, як саме ми представляємо сигнал у вигляді розкладу по деякому базису $\{\varphi_i\}_{i=1}^N$. Використаємо поняття скалярного добутку. Помножимо скалярно обидві частини рівняння (1.17) справа на один з векторів базису φ_j :

$$\langle x, \varphi_j \rangle = \sum_{i=1}^N c_i \langle \varphi_i, \varphi_j \rangle = c_1 \langle \varphi_1, \varphi_j \rangle + c_2 \langle \varphi_2, \varphi_j \rangle + \dots + c_N \langle \varphi_N, \varphi_j \rangle.$$

В цьому виразі відомі вектор x та всі вектори $\{\varphi_i\}_{i=1}^N$, отже, отримано одне рівняння, в якому невідомими є тільки всі коефіцієнти $\{c_i\}$. Повторимо цю процедуру для всіх елементів базису $j=1,2,3\dots N$, і отримаємо систему N лінійних скалярних рівнянь відносно всіх a_i , яку можна записати в матричному вигляді:

$$\begin{bmatrix} \langle \varphi_1, \varphi_1 \rangle & \langle \varphi_1, \varphi_2 \rangle & \dots & \langle \varphi_1, \varphi_N \rangle \\ \langle \varphi_2, \varphi_1 \rangle & \langle \varphi_2, \varphi_2 \rangle & \dots & \langle \varphi_2, \varphi_N \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \varphi_N, \varphi_1 \rangle & \langle \varphi_N, \varphi_2 \rangle & \dots & \langle \varphi_N, \varphi_N \rangle \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} \langle x, \varphi_1 \rangle \\ \langle x, \varphi_2 \rangle \\ \vdots \\ \langle x, \varphi_N \rangle \end{bmatrix}, \quad (1.18)$$

або $\Phi c = \beta$. Тут Φ – матриця системи з відомими елементами (матриця Грама), c – вектор-стовпець невідомих коефіцієнтів розкладу, які треба знайти, β – вектор-стовпець вільних членів.

Розв'язок цієї системи дає нам вектор невідомих коефіцієнтів $c = [c_1, c_2, \dots, c_N]^T$, які є координатами сигналу x в просторі M відносно базису $\{\varphi_i\}_{i=1}^N$:

$$c = \Phi^{-1} \beta.$$

Ця неоднорідна система лінійних рівнянь буде мати розв'язок, якщо визначник матриці коефіцієнтів системи не дорівнює нулю. Цей визначник називається визначником Грама. Можна довести, що цей визначник не рівний нулю, оскільки він дорівнює об'єму гіперпараллелограма, що побудований на базисних векторах в просторі відповідної розмірності.

Для знаходження вектору невідомих коефіцієнтів c можна скористатися ще одним методом. Введемо (оберемо або задамо) в просторі M_N ще один базис $\{v_i\}_{i=1}^N$, всі вектори якого є попарно ортогональні всім векторам базису $\{\varphi_i\}_{i=1}^N$:

$$\langle \varphi_i, v_j \rangle = \begin{cases} r, & i = j, \\ 0, & i \neq j. \end{cases} \quad (1.19)$$

Помножимо скалярно обидві частини (1.17) на довільний вектор v_j з базису $\{v_i\}_{i=1}^N$

$$\langle x, v_j \rangle = \sum_{i=1}^n c_i \langle \varphi_i, v_j \rangle.$$

В правій частині рівності стоїть сума скалярних добутків векторів з базису $\{\varphi_i\}_{i=1}^N$ на вектор v_j . З урахуванням того, що базиси ортогональні один одному, з усієї суми ненульовим виявиться лише один доданок, той, для якого співпадають номери векторів $i = j$:

$$\langle x, v_j \rangle = \sum_{i=1}^n c_i \langle \varphi_i, v_j \rangle = c_j \cdot r.$$

Отже, коефіцієнт розкладу c_j вектора x в базисі $\{\varphi_i\}_{i=1}^N$ можна отримати як скалярний добуток цього вектора на вектор v_j . Повторюючи цю процедуру для всіх векторів з допоміжного базису $\{v_i\}_{i=1}^N$, можна виразити значення всіх невідомих коефіцієнтів r_j :

$$c_j = \frac{\langle x, v_j \rangle}{r}.$$

Часто вимагають, щоб вектори в базисах були ортонормованими, тобто $r = 1$. Базис, який задовольняє умовам (1.19) для даного базису, називається *взаємним базисом* (reciprocal basis). Отже, для таких взаємних базисів ми маємо вираз для розкладу довільного вектора x в одному з базисів через скалярні добутки з векторами іншого базису. Причому, якщо поміняти місцями базиси $\{v_i\}_{i=1}^N$ та $\{\varphi_i\}_{i=1}^N$, всі викладки залишаться справедливими:

$$x = \sum_{i=1}^N \langle x, v_i \rangle \varphi_i = \sum_{i=1}^N \langle x, \varphi_i \rangle v_i.$$

Скалярні коефіцієнти $\{c_i\}$ розкладу в базисі $\{\varphi_i\}_{i=1}^N$ вектора x можна розглядати як N -вимірний вектор в просторі впорядкованих сукупностей дійсних чисел R^N або комплексних чисел C^N . Вектор x може вважатися математичною моделлю деякого сигналу, причому властивості сигналу та векторів, які їх моделюють, пов'язані.

Набір з N чисел $\{c_i\}$ є коефіцієнтами, які дають представлення вектора x у просторі M_N відносно базису $\{\varphi_i\}_{i=1}^N$. Отже, існує однозначна відповідність між довільними векторами в просторі M_N та простором N -вимірних векторів-рядків чисел, тобто між сигналами та наборами коефіцієнтів їх розкладу в деякому базисі.

Треба зауважити, що представлення не мають змісту самі по собі, а обов'язково мають бути віднесені до деякого базису. Також треба мати на увазі, що деякі математичні перетворення сигналів можна розглядати як їх представлення в тому чи іншому просторі, або перехід від представлення в одному просторі до представлення того ж сигналу в іншому просторі. Властивості базису будуть впливати на представлення в ньому сигналів.

Розглянемо окремий випадок розкладу сигналу в базисі, вектори в якому попарно ортогональні, тобто

$$\langle \varphi_i, \varphi_j \rangle = \begin{cases} \|\varphi_i\|^2, & i = j, \\ 0, & i \neq j. \end{cases} \quad (1.20)$$

При скалярному множенні лівої і правої частин цього рівняння на довільний вектор з базису, з урахуванням ортогональності системи $\{\varphi_i\}_{i=1}^N$ виявляється, що всі доданки перетворюються на нуль, окрім єдиного, для якого отримаємо

$$\langle x, \varphi_i \rangle = c_i \|\varphi_i\|^2.$$

В цьому випадку матриця попарних скалярних добутків Φ в (1.18) стає діагональною матрицею (говорять, що ортогональний базис діагоналізує матрицю) – ненульові елементи залишаються лише на головній діагоналі. Визначник матриці Грама в цьому випадку буде рівний

$$G = \prod_{i=1}^N \|\varphi_i\|^2.$$

В такому випадку одразу можна отримати явний вираз для розрахунку коефіцієнтів, кожний з яких дорівнює значенню проєкції сигналу на відповідний базисний вектор:

$$c_i = \frac{\langle x, \varphi_i \rangle}{\|\varphi_i\|^2}. \quad (1.21)$$

Представлення сигналу у вигляді лінійної комбінації базисних векторів в ортогональному базисі з коефіцієнтами (1.21) набуває вигляду:

$$x = \sum_{i=1}^N \frac{\langle x, \varphi_i \rangle}{\|\varphi_i\|^2} \varphi_i. \quad (1.22)$$

Якщо базис $\{\varphi_i\}_{i=1}^N$ є ортонормованим, то процедура обчислення коефіцієнтів розкладу в ньому ще більше спрощується. *Ортонормований* базис – це такий базис, для якого виконуються співвідношення:

$$\langle \varphi_i, \varphi_j \rangle = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

В ортонормованому базисі представлення сигналу матиме вигляд:

$$x = \sum_{i=1}^{\infty} \langle x, \varphi_i \rangle \varphi_i. \quad (1.23)$$

В цьому випадку коефіцієнти розкладу сигналу дорівнюють $c_i = \langle x, \varphi_i \rangle$.

Рівність Парсеваля

Нехай маємо ортогональний базис. Знайдемо квадрат норми вектора x , який представлений в цьому базисі. Норма для векторів, що відповідають сигналам електронних систем має енергетичний зміст, а саме, квадрат норми дорівнює енергії сигналу.

$$\begin{aligned} \|x\|^2 &= \langle x, x \rangle = \left\langle \sum_{i=1}^N c_i \varphi_i, \sum_{i=1}^N c_i \varphi_i \right\rangle = \\ &= \langle c_1 \varphi_1 + c_2 \varphi_2 + \dots, c_1 \varphi_1 + c_2 \varphi_2 + \dots \rangle = \\ &= \langle c_1 \varphi_1, c_1 \varphi_1 \rangle + \langle c_2 \varphi_2, c_2 \varphi_2 \rangle + \dots + \langle c_1 \varphi_1, c_2 \varphi_2 \rangle + \langle c_1 \varphi_1, c_3 \varphi_3 \rangle + \dots = \\ &= |c_1|^2 \langle \varphi_1, \varphi_1 \rangle + |c_2|^2 \langle \varphi_2, \varphi_2 \rangle + |c_3|^2 \langle \varphi_3, \varphi_3 \rangle + \dots + |c_N|^2 \langle \varphi_N, \varphi_N \rangle \end{aligned}$$

Було застосовано властивість, що скалярні добутки двох різних базисних ортогональних векторів дорівнюють нулю. Отже, остаточно можна записати, що

$$\|x\|^2 = \sum_{i=1}^N |c_i|^2 \|\varphi_i\|^2,$$

а для ортонормованої системи

$$\|x\|^2 = \sum_{i=1}^N |c_i|^2.$$

Отже, квадрат норми сигналу дорівнює сумі квадратів координат цього вектора відносно ортонормованого базису. Це твердження називається *рівністю Парсеваля*. Воно говорить, що енергія сигналу складається з енергій його спектральних компонент.

Повнота систем ортонормованих функцій

Розглянемо систему ортонормованих функцій $\{\varphi_i(t)\}_{i=1}^{\infty}$, яка не обов'язково є базисом. Нехай для будь-якого номера k та довільної функції $f(t)$ виконується твердження:

$$\int \varphi_k(t) f(t) dt = 0 \Rightarrow f(t) = 0$$

майже всюди в області визначення функцій $\varphi_i(t)$ та $f(t)$. Це твердження говорить про те, що в просторі не існує більше жодної функції, яка була б ортонормованою до тих функцій, які вже є в системі $\{\varphi_i(t)\}_{i=1}^{\infty}$. При цьому система $\{\varphi_i(t)\}_{i=1}^{\infty}$ називається *повною системою функцій*. З повноти системи функцій випливає, що це є базис для відповідного простору.

Нерівність Бесселя

Розглянемо *ортонормовану систему функцій* $\{\varphi_i(t)\}_{i=1}^{\infty}$. Нехай також маємо деяку функцію $f(t)$, яка належить тому самому простору, в якому задана ортонормована система. Поставимо у відповідність цій функції суму, яка є лінійною комбінацією базисних функцій (в ній використано лише N доданків):

$$f(t) \rightarrow S_N = \sum_{i=1}^N c_i \varphi_i(t).$$

Задамо питання, при яких коефіцієнтах c_i відстань між функцією $f(t)$ і S_N буде мінімальною. Відстань визначимо за допомогою метрики:

$$D_2(f(t), S_N) = \sqrt{\int \left(f(t) - \sum_{i=1}^N c_i \varphi_i(t) \right)^2 dt} \rightarrow \min$$

Ця відстань буде сягати свого мінімуму, коли мінімальним буде підкореневий вираз. Виконаємо деякі перетворення:

$$\begin{aligned} D_2^2(f(t), S_N) &= \int \left(f(t) - \sum_{i=1}^N c_i \varphi_i(t) \right)^2 dt = \\ &= \int f^2(t) dt - 2 \int f(t) \sum_{i=1}^N c_i \varphi_i(t) dt + \int \left(\sum_{i=1}^N c_i \varphi_i(t) \right)^2 dt. \end{aligned}$$

Для знаходження коефіцієнтів c_i , при яких відстань буде мінімальною, треба знайти похідні цієї функції по кожному з коефіцієнтів, та прирівняти їх до нуля:

$$\frac{d}{dc_i} D_2^2(f(t), S_N) = 0 \Rightarrow c_i.$$

Оскільки перший інтеграл в попередньому виразі не залежить від c_i , похідна від нього буде рівною нулю:

$$\frac{dI_1}{dc_i} = 0.$$

Розглянемо по черзі похідні від другого та третього інтегралів.

В другому інтегралі поміняємо порядок інтегрування і додавання та розпишемо інтеграл як скалярний добуток:

$$\begin{aligned} \frac{dI_2}{dc_i} &= \frac{d}{dc_i} \left(-2 \sum_{i=1}^N c_i \int f(t) \varphi_i(t) dt \right) = -2 \frac{d}{dc_i} \left(\sum_{i=1}^N c_i \langle f(t), \varphi_i(t) \rangle \right) = \\ &= -2 \frac{d}{dc_i} (c_1 \langle f(t), \varphi_1(t) \rangle + c_2 \langle f(t), \varphi_2(t) \rangle + \dots \\ &\quad + c_i \langle f(t), \varphi_i(t) \rangle + \dots + c_N \langle f(t), \varphi_N(t) \rangle) = \\ &= \frac{d}{dc_i} (-2 c_i \langle f(t), \varphi_i(t) \rangle) = -2 \langle f(t), \varphi_i(t) \rangle. \end{aligned}$$

Видно, що в сумі скалярних добутків коефіцієнт c_i , по якому береться похідня, присутній лише в одному доданку. Отже, похідні по c_i від всіх інших доданків в попередній сумі будуть рівними нулю, оскільки $k \neq i$.

В останньому інтегралі розкриємо дужки для квадрату суми:

$$\begin{aligned} I_3 &= \int \left(\sum_{i=1}^N c_i \varphi_i(t) \right)^2 dt = \\ &= \int (c_1 \varphi_1(t) + c_2 \varphi_2(t) + \dots + c_i \varphi_i(t) + \dots + c_N \varphi_N(t)) \times \\ &\quad \times (c_1 \varphi_1(t) + c_2 \varphi_2(t) + \dots + c_i \varphi_i(t) + \dots + c_N \varphi_N(t)) dt = \\ &= \sum_{i \neq k} \int c_i c_k \varphi_i(t) \varphi_k(t) dt + \sum_{i=1}^N \int c_i^2 \varphi_i^2(t) dt. \end{aligned}$$

В першій сумі кожний інтеграл є скалярним добутком двох різних ортогональних функцій, отже він рівний нулю. В другій сумі буде лише один доданок, в який увійде c_i , який рівний $\int c_i^2 \varphi_i^2(t) dt = c_i^2 \|\varphi_i(t)\| = c_i^2$ оскільки базис є ортонормований.

Отже,

$$\frac{dI_3}{dc_i} = 2c_i.$$

Разом будемо мати

$$\frac{d}{dc_i} D_2^2(f(t), S_N) = -2\langle f(t), \varphi_i(t) \rangle + 2c_i = 0$$

$$c_i = \langle f(t), \varphi_i(t) \rangle.$$

Отже, видно, що саме при таких коефіцієнтах c_i функція, яка представляє лінійну комбінацію ортонормованих векторів $\{\varphi_i(t)\}_{i=1}^N$ буде *найближчою* до нашого початкового сигналу.

Коефіцієнти $c_i = \langle f(t), \varphi_i(t) \rangle$ називаються коефіцієнтами Фур'є, а розклад нашої функції у ряд з такими коефіцієнтами називається *узагальненим рядом Фур'є*:

$$f(t) = \sum_{i=1}^N c_i \varphi_i(t).$$

Той факт, що при таких коефіцієнтах різниця $D_2(f(t), S_N)$ буде мінімальною, називається *екстремальною властивістю* ряду Фур'є.

Знайдемо цю мінімальну різницю:

$$\begin{aligned} D_{2\min}^2(f(t), S_N) &= \int f^2(t) dt - 2 \int f(t) \sum_{i=1}^N c_i \varphi_i(t) dt + \int \left(\sum_{i=1}^N c_i \varphi_i(t) \right)^2 dt = \\ &= \|f(t)\|^2 - 2 \sum_{i=1}^N c_i \int f(t) \varphi_i(t) dt + \sum_{i=1}^N c_i^2 \int \varphi_i^2(t) dt = \\ &= \|f(t)\|^2 - 2 \sum_{i=1}^N c_i^2 + \sum_{i=1}^N c_i^2 = \|f(t)\|^2 - \sum_{i=1}^N c_i^2. \end{aligned}$$

Оскільки попередній вираз – квадрат метрики, то він невід'ємний, отже, остаточно можна записати

$$\|f(t)\|^2 \geq \sum_{i=1}^N c_i^2. \quad (1.24)$$

Отриманий вираз називається *нерівність Бесселя*.

Ортонормована система функцій називається *повною*, якщо для неї нерівність Бесселя перетворюється на рівність Парсеваля. Отже, набір функцій є базисом. В цьому випадку енергія сигналу рівна сумі енергій складових розкладу.

В нескінченновимірному просторі норма похибки може бути зроблена як завгодно малою шляхом додавання до базису все більшої кількості ортогональних векторів. З практичної точки зору таке наближення є корисним, оскільки існує дуже багато різних повних ортонормованих систем, якими можна користуватися. Також зручним є те, що внаслідок ортонормованості, якщо ми додаємо до набору $\{\varphi_i(t)\}_{i=1}^N$ ще один ортогональний вектор $\varphi_{N+1}(t)$, щоб уточнити апроксимацію, нам немає необхідності

перераховувати заново всі коефіцієнти розкладу, а достатньо розрахувати тільки $\langle f(t), \varphi_{N+1}(t) \rangle$. Це пов'язано з тим, що ортонормований базис є самоспряженим, тобто він співпадає з взаємним базисом. Тому в основному надалі ми будемо розглядати різні ортонормовані системи базисних функцій.

Але постає питання, як саме треба обирати цей підпростір, тобто, які лінійно-незалежні функції обирати для представлення сигналу. Задача знаходження оптимального підпростору представлення сигналу має сенс лише для деяких обмежених підпросторів сигналів, які відносяться до певного виду. Ця задача пов'язана також з задачею побудови адаптивних базисів розкладу, для яких властивості базисних функцій якимось чином пов'язані з властивостями сигналів, або які побудовані саме для того, щоб представлення конкретного сигналу в таких базисах виявляло в сигналах потрібні нам особливості.

1.8. Теорема ортогонального проектування

В попередньому підрозділі було розглянуто випадок, коли вектор x , який треба було розкласти в базисі $\{\varphi_i\}_{i=1}^N$ підпростору M_N , належав цьому підпростору: $x \in M_N$. В цьому випадку його можна розкласти в базисі $\{\varphi_i\}_{i=1}^N$, і рівність (1.17) буде виконуватися точно.

Розглянемо тепер інший випадок, коли вектор x не належить підпростору M_N , $x \notin M_N$ тобто не може бути представлений у вигляді лінійної комбінації базисних векторів. В такому випадку для представлення вектору x в цьому базисі необхідно знайти близький до нього вектор з простору M_N . Мірою близькості може бути відстань від нашого вектора до вектора, який розкладається в базисі точно.

Нехай простір M_N – підпростір гільбертового простору $M_N \in L^2$. Ми можемо поставити у відповідність кожному довільному вектору $x \in L^2$ деякий вектор \hat{x} з підпростору M_N , який найбільш близький до вектора x . Можна довести таку теорему:

Теорема проектування:

Для довільного вектора з лінійного N -вимірного простору $x \in L^2$ існує єдиний вектор $\hat{x} \in M_N$, який задається розкладом

$$\hat{x} = \sum_{i=1}^N \langle x, \varphi_i \rangle \varphi_i = \sum_{i=1}^N c_i \varphi_i, \quad (1.25)$$

такий, що різниця $\rho = x - \hat{x}$ ортогональна всім векторам в M_N , та $\|x - \hat{x}\| < \|x - \tilde{x}\|$, де \tilde{x} – будь-який інший вектор з M_N .

Вектор \hat{x} називається *ортогональною проекцією* вектора x на підпростір M_N , або апроксимацією вектора x за допомогою базису $\{\varphi_i\}_{i=1}^N$, а вектор $\rho = x - \hat{x}$ – *похибкою наближення* вектора x вектором \hat{x} .

Різниця $\rho = x - \hat{x}$ між тим вектором x , який ми хочемо представити в базисі, і тим, який буде найбільш близьким до нього з лінійної оболонки базисних векторів, буде ортогональною всім векторам базису $\{\varphi_i\}_{i=1}^N$. Тому вона не може бути представлена за допомогою лінійної комбінації базисних векторів, і завжди залишається ненульовою.

Точність наближення вектора характеризується нормою похибки:

$$\|\rho\|^2 = \|x\|^2 - \|\hat{x}\|^2. \quad (1.26)$$

Отже, рівність $x = \sum_{i=1}^N c_i \varphi_i$ буде точною лише у випадку, коли сигнал x належить до множини лінійних комбінацій базисних векторів $\{\varphi_i\}_{i=1}^N$. Якщо це не так, то лінійна комбінація базисних векторів з коефіцієнтами $\{c_i\}$, дає деякий сигнал \hat{x} , який є найкращим **наближенням** до сигналу x , яке можна представити у базисі $\{\varphi_i\}_{i=1}^N$.

Наближення є найкращим в тому значенні, що воно забезпечує мінімально можливу норму похибки представлення сигналу:

$$\|x - \hat{x}\| = \left\| x - \sum_{i=1}^N c_i \varphi_i \right\| \rightarrow \min \text{ при } N \rightarrow \infty.$$

Можна довести за допомогою апарату функціонального аналізу, що простір L^2 є повним сепарабельним простором, тобто ортогональним проектуванням можна отримати як завгодно близьку апроксимацію будь-якого сигналу $x \in L^2$, якщо обрати кількість векторів в базисі достатньо великою.

Отже, найкращим представленням сигналу в лінійному нормованому просторі з метрикою, що породжена нормою, і з нормою, породженою скалярним добутком в гільбертовому просторі, є представлення, при якому норма різниці між сигналом і вектором, який його представляє, є мінімальною. Пошук такого представлення полягає у ортогональному проектуванні сигналу на підпростір, натягнутий на базис цього простору. Коефіцієнти представлення можуть бути знайдені як скалярні добутки сигналу на відповідні базисні функції.

1.9. Розклади зі збереженням енергії

Співвідношення $\|f(t)\|^2 \geq \sum_{i=1}^N c_i^2$ для сигналів, які є функціями часу, набуває енергетичного змісту. Якщо під сигналом мається на увазі електричне коливання (струм або напруга), то величина

$$\|f(t)\|^2 = \int_{t_1}^{t_2} |f(t)|^2 dt = E$$

є енергією сигналу в проміжку часу від t_1 до t_2 , за умови, що опір, в якому виділяється енергія, дорівнює 1 Ом.

В такому разі енергія сигналу дорівнює в загальному випадку

$$E = \sum_{i=1}^{\infty} c_i^2 \|\varphi_i(t)\|^2,$$

а для представлення сигналу в системі повних ортонормованих функцій

$$E = \sum_{i=0}^{\infty} c_i^2.$$

Рівність Парсеваля в цьому випадку говорить про те, що розклад сигналу за деякою системою функцій дає його представлення *зі збереженням енергії*. Це означає, що енергія сигналу при його представленні в базисі не втрачається, а лише перерозподіляється між елементарними складовими розкладу (базисними функціями), якими він представлений. Енергетичні характеристики сигналу, розраховані в часовій області, та ті, які в сукупності мають всі складові розкладу – однакові.

1.10. Види та характеристики біомедичних сигналів

Існування живих організмів, в першу чергу людини, супроводжується великою кількістю сигналів. Ці сигнали мають різну природу та пов'язані із органами та системами організму.

Класифікація біосигналів може бути проведена за багатьма критеріями.

За природою виникнення:

- електричні;
- магнітні;
- хімічні;
- механічні;

- акустичні;
- оптичні;
- термічні.

Відповідно до системи організму, де виникають сигнали:

- ендокринна;
- нервова (центральна та периферична);
- серцево-судинна;
- система зору;
- система слуху;
- опорно-руховий апарат;
- система дихання;
- шлунково-кишковий тракт;
- система крові.

Електричні сигнали – це струми, які протікають в тілі людини, та напруги, які виникають на поверхні тіла через роботу органів. Електричні поля генеруються клітинами (нервовими та м'язовими) через протікання позаклітинних та внутрішньоклітинних іонних струмів.

Види електричних сигналів:

- електронейрограма;
- електрокардіограма;
- електроенцефалограма;
- електроміограма;
- електрогастрограма;
- електроретінограма;
- електроокулограма.

Магнітні сигнали:

- магнітнейрограма;
- магнітоенцефалограма;
- магнітокардіограма;
- магнітоміограма.

Хімічні сигнали – це зареєстровані зміни концентрацій певних речовин в часі.

Види:

- рівень глюкози в крові;
- рівень кисню та інших газів в крові:
 - SaO_2 – сатурація артеріальної крові киснем – відсоток молекул гемоглобіну, які тримають молекулу кисня;
 - SvO_2 – сатурація венозної крові киснем, показує, скільки кисню споживає тіло людини;

- SpO_2 – концентрація кисню в периферичних капілярах, це те саме, що і SaO_2 , але для капілярів;
- $SpCO_2$ – концентрація карбоксигемоглобіну в крові. Цей сигнал вимірюється з фотоплетизмограми.
- рівень концентрації газів в повітрі, що видихається;
- рН.

Механічні біосигнали – відображають механічні функції тіла:

- тиск крові;
- рухи людини (прискорення та зміщення);
- рухи частин тіла людини (грудна клітина, кінцівки та ін.);
- характеристики потоку повітря під час дихання.

Акустичні біосигнали – звуки людини:

- голос (вокалограма);
- звуки серця (фонокардіограма);
- хропіння (детектування апное);
- звуки дихання
- ковтання;
- звуки в суглобах.

Оптичні біосигнали – випромінювання під час реакцій, зокрема при генетичних та клітинних дослідженнях.

- Випромінювання організму
- Викликані сигнали (відбиття та викликана флуорисценція)

Термічні біосигнали:

- Термограма (карта розподілу температур)

1.11. Висновки до розділу 1 та рекомендована література

В цьому розділі було розглянуто види сигналів та способи їх представлення. Визначено, що сигнали можна класифікувати за багатьма різними ознаками. Найбільш суттєвим є розділення на неперервні та дискретні сигнали: в реальному житті всі сигнали переважно неперервні (аналогові), але майже всі технічні засоби працюють лише з дискретним представленням сигналів. Через це важливою операцією є аналого-цифрове перетворення: заміна неперервних сигналів їх відліками, виміряними в певні моменти часу. Для того, щоб адекватно представити початковий неперервний сигнал в цифровій системі обробки та аналізу, виконується

не тільки дискретизація сигналу в часі, але також квантування отриманих відліків за величиною і подальше кодування отриманих чисел в певною роздільною здатністю. Хоча всі сигнали можна описати неперервними функціями або числовими послідовностями, існує більш компактне представлення сигналів – як елементів деякого простору. В просторі сигналів кожен з сигналів – це точка, положення якої задається координатами. Як правило, це гільбертів простір сигналів: лінійний простір, в якому можна вимірювати відстані між сигналами, для кожного сигналу можна визначити його розмір (норму), та в якому визначена операція скалярного добутку двох сигналів. В просторах сигналів існує багато базисів з різними властивостями, але у всіх цих базисах можна представляти сигнали, що підлягають аналізу, і отримувати таким чином розклади сигналів на більш прості складові.

В залежності від властивостей сигналів і базисних функцій, такі представлення можуть бути точними або давати тільки апроксимацію сигналу з деякою точністю. Якщо базис в просторі сигналів є ортогональним, то розклад буде проводитись зі збереженням енергії, що дозволяє переходити від сигналу до представлення і навпаки без втрати інформації про властивості сигналу; внаслідок цього можна аналізувати і самі сигнали, і їх представлення.

Класифікація сигналів та їх властивості добре висвітлені в [1-5], тут також можна прочитати про історію теорії сигналів та приклади її застосувань. Докладний опис дискретизації сигналів та деякі питання аналого-цифрового перетворення можна знайти в [2, 6-9]. Для більш поглибленого вивчення питань представлення сигналів в базисах та розкладів сигналів, а також для повторення основних понять лінійної алгебри, варто звернутися до [10-13]. Докладний розгляд та більш глибоке висвітлення питань розкладів сигналів представлений в [14-15]. Цікавому питанню побудови адаптивних базисів розкладу та методів пошуку оптимальних представлень сигналів присвячені роботи [16-18].

1.12. Питання для самоперевірки

1. Що таке сигнал, за якими ознаками їх можна класифікувати? Чим відрізняються неперервні та дискретні сигнали, якими функціями вони моделюються?
2. Як відбувається аналого-цифрове перетворення сигналів, з якою метою воно виконується? Які основні характеристики АЦП?
3. Що таке цифрова послідовність, які математичні операції можна над ними виконувати? Як пов'язані між собою дискретні сигнали і числові послідовності?
4. Якими бувають лінійні простори сигналів? Навіщо використовують скалярний добуток, норму, метрику; за якими виразами їх розраховують?
5. В чому полягає розклад сигналу в базисі, як отримати коефіцієнти розкладу для випадку ортогональних та неортогональних базисів?

6. Коли сигнал можна представити в базисі точно, а коли можна говорити лише про наближення сигналу за допомогою базисних функцій?
7. В чому полягає властивість збереження енергії при розкладі в ортогональних базисах? Які наслідки має ця властивість з точки зору користі для аналізу сигналів?

2. ДИСКРЕТНІ СИСТЕМИ

2.1. Загальні відомості про системи обробки сигналів

Equation Chapter 1 Section 2

Сигнали та системи, які призначені для їх аналізу, нерозривно пов'язані. Не буває сигналу окремо від тієї системи, яка створена для того, щоб зареєструвати, підсилити сигнал, обробити його, зберегти, проаналізувати або передати іншій системі. Так само, не може існувати просто система сама по собі, окремо від тих сигналів, які дослідження яких вона створена.

Через це неможливо вивчати теорію сигналів без знання того, які саме системи призначені для обробки та аналізу сигналів, які властивості вони мають та які можливості надають дослідникам. Вся обробка та аналіз сигналів відбувається лише за допомогою систем, і від того, наскільки правильно буде спроектована система обробки сигналу, буде залежати те, наскільки правильний результат буде отримано в результаті.

Система – сукупність елементів, що перебувають у відношеннях та зв'язках один з одним, яка утворює певну цілісність та єдність.

Будь-яка система повинна підкорятися таким системним принципам:

цілісність – не можна звести властивості системи до суми властивостей її складових, неможливо вивести її властивості з суми властивостей складових. Система є єдиним цілим, і тільки у взаємодії всіх складових частин виявляє свої властивості;

структурність – можливість описання системи через встановлення її структури (мережі зв'язків та відношень елементів), залежність властивостей системи не тільки від властивостей та поведінки елементів, а й від структури системи;

взаємозалежність системи і середовища – система формує і проявляє свої властивості в процесі взаємодії з зовнішнім для неї середовищем;

ієрархічність – кожний компонент системи може розглядатися як система, а сама система є компонентом більшої системи.

Для вивчення теорії сигналів цікавими є такі чотири види систем залежно від призначення:

- системи генерації сигналів;
- системи обробки сигналів;
- системи управління;
- системи передачі сигналів.

Системи цифрової обробки сигналів.

Блок-схема цифрової обробки сигналів (ЦОС) наведена на рис. 2.1.

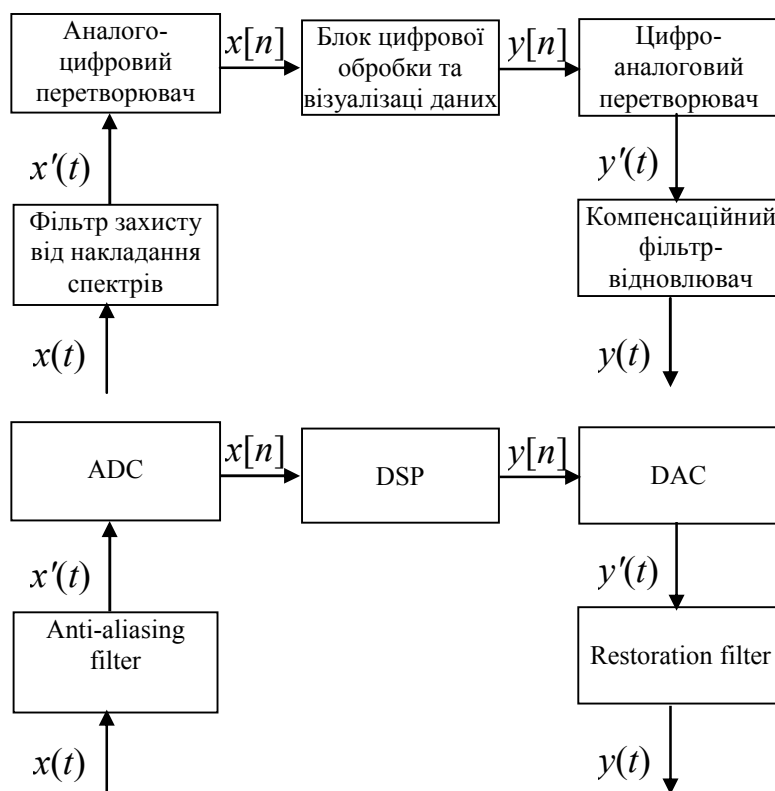


Рисунок 2.1 – Блок-схема системи ЦОС

Вхідний аналоговий сигнал $x(t)$ спочатку поступає на фільтр нижніх частот, призначений для захисту від явища накладання спектрів та зменшення спотворень в сигналі після дискретизації. Більш докладно про призначення та необхідність цього фільтру буде вестися мова в розділі, присвяченому дискретизації сигналів та частотному представленню дискретизації.

Після проходження через вхідний фільтр, відфільтрований вхідний аналоговий сигнал $x'(t)$ в блоці аналого-цифрового перетворювача (АЦП) піддається дискретизації та квантуванню. На виході АЦП утворюється послідовність цифрових відліків вхідного сигналу $x[n]$, яка поступає у блок цифрової обробки та аналізу сигналу. В цьому блоці реалізований потрібний досліднику алгоритм перетворення сигналів для отримання необхідних параметрів сигналу та їх візуалізації. В системах, які призначені суто для аналізу сигналів, цей блок є останнім. В ньому будуть розраховані та проаналізовані необхідні параметри вхідного сигналу.

Якщо за результатами обробки вхідного сигналу потрібне не лише визначення параметрів сигналу, а треба згенерувати вихідний сигнал, який модифікований відносно вхідного, то блок цифрової обробки формує послідовність вихідних цифрових відліків $y[n]$. Цей дискретний сигнал має інші властивості відносно вхідного сигналу $x[n]$ (це може бути підсилений сигнал, відфільтрований та т.і.)

Вихідний дискретний сигнал потім надходить до цифро-аналогового перетворювача (ЦАП), який перетворює вхідний цифровий код, що відповідає значенням сигналу

$y[n]$ в кожний момент часу, у вихідний кусково-неперервний сигнал $y'(t)$ (квантований за величиною та неперервний у часі). Після проходження компенсаційного фільтра-відновлювача, призначеного для згладжування та зменшення спотворень після цифро-аналогового перетворення, на виході всієї системи ЦОС з'являється вихідний неперервний сигнал $y(t)$.

Систему ЦОС на рис. 2.1 можна назвати цифровою системою обробки аналогових сигналів, оскільки і вхідний, і вихідний сигнали цієї системи є неперервними, а потрібний метод обробки сигналів реалізований у вигляді алгоритму, що працює з цифровим сигналом. Щоб перейти від вхідного неперервного сигналу до його представлення у цифровому вигляді, а потім ще раз до вихідного неперервного сигналу – в системі присутні допоміжні блоки фільтрів, АЦП та ЦАП.

Сутність цифрової обробки полягає в тому, що вхідний аналоговий сигнал після попередньої фільтрації перетворюється у послідовність чисел, яка потім піддається математичним перетворенням у обчислювальному пристрої. В залежності від складності методу обробки, обчислення можуть бути реалізовані на мікроконтролерах, цифрових сигнальних процесорах або на комп'ютері. За потреби трансформований цифровий сигнал (послідовність чисел) може бути перетворена знову на неперервний сигнал за допомогою ЦАП.

До переваг цифрової обробки сигналів відносяться:

- гарантована точність, яка визначається лише кількістю бітів, задіяних для представлення відліків сигналу;
- повна повторюваність сигналу та результатів обробки, якщо алгоритми обробки ідентичні;
- відсутність спотворення внаслідок старіння апаратури та впливів факторів навколишнього середовища;
- підвищена надійність, низька вартість, малі розміри, низьке енергоспоживання;
- гнучкість: системи ЦОС можна багато разів перепрограмувати на виконання різних функцій без зміни обладнання;
- значна продуктивність та можливість реалізувати функції, неможливі в аналогових системах, складні алгоритми перетворень сигналів.

Недоліки цифрової обробки сигналів:

- існування похибок внаслідок скінченної розрядності представлення чисел;
- порівняно низька швидкість роботи, особливо для широкосмугових сигналів та надвисокочастотних сигналів;
- значні витрати часу на розробку цифрових систем;
- відносна складність побудови систем ЦОС, потреба у великій кількості допоміжних блоків.

Переваги аналогової обробки сигналів:

- робота в реальному часі, незамінність для надшвидких сигналів і для задач спеціальної обробки суто аналогових сигналів;
- порівняна простота електронних схем;
- малі порівняно з ЦОС економічні витрати.

Недоліки аналогової обробки сигналів:

- невелика точність перетворення;
- неможливість або складність зміни алгоритмів перетворення сигналу в аналоговій електронній схемі;
- залежність показників від зовнішніх факторів кліматички та живлення;
- нагромадження шумів різної природи.

Системи зв'язку

Системи зв'язку призначені для передачі інформації за допомогою сигналів від джерела повідомлення до споживача. Класична функціональна схема системи зв'язку була запропонована Клодом Елвудом Шенноном в 1948 році в роботі «Математична теорія зв'язку» (рис. 2.2).

Інформація, яку треба передати, представляється у вигляді повідомлення. Повідомлення, яке поступає від джерела, обробляється в передавачі, в якому формується сигнал, зручний для передачі в лінії зв'язку. Наприклад, в телефонії звуковий тиск перетворюється у пропорційний змінний електричний струм мікрофона. За допомогою передавача сигнал, що несе повідомлення, відправляється в лінію зв'язку. Лінія зв'язку – середовище, яке використовуються для передачі сигналів від передавача до приймача. Це може бути пара дротів, кабель, хвилевод, або простір, заповнений повітрям. Після проходження лінії зв'язку сигнал потрапляє до приймача. Приймач реєструє вхідний сигнал та виділяє з нього повідомлення, яке передається споживачу. Всі перешкоди, які взаємодіють з сигналом під час проходження в лінії зв'язку, спотворюючи сигнал та заважаючи передачі, моделюються в запропонованій Шенноном схемі за допомогою джерела перешкод. Передавач, приймач та лінія зв'язку разом з джерелом перешкод називаються *каналом зв'язку*.

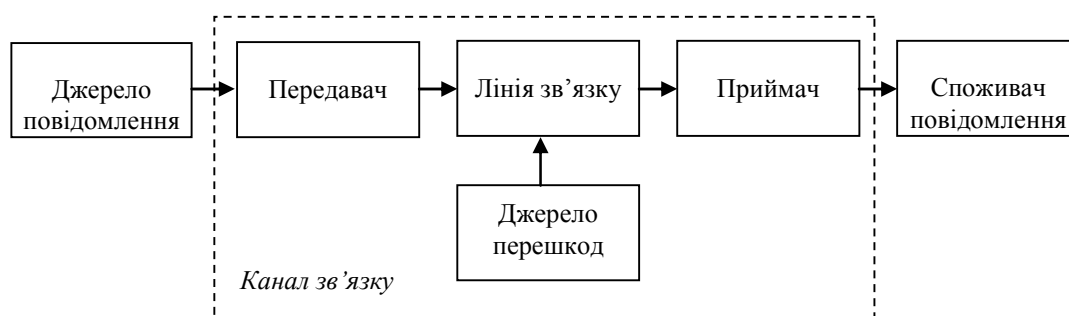


Рисунок 2.2 – Функціональна схема системи зв'язку

Системи управління

Управління – це цілеспрямований вплив на об'єкт або процес, що веде до заданої зміни його стану або до підтримки його в заданому стані.

Блок-схема системи управління наведена на рис. 2.3.

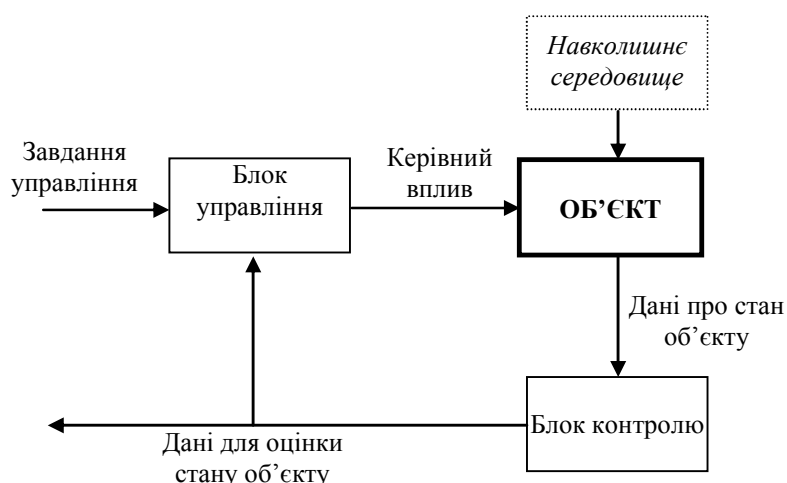


Рисунок 2.3 – Блок-схема системи управління

На основі завдання, яке ставиться для управління об'єктом для досягнення певної мети, блок управління генерує сигнали керівного впливу на об'єкт. Одночасно блок контролю за об'єктом за допомогою датчиків постійно збирає інформацію про поточний стан об'єкту. Зареєстровані від об'єкта сигнали постійно обробляються в реальному часі, та на виході блоку контролю генерується інформація про стан об'єкта. Вона передається на блок управління, в якому поточний стан об'єкта порівнюється з тим станом, якого треба досягти відповідно до завдання управління. Відповідно до результатів такого порівняння генерується або корегується керівний вплив на об'єкт. Залежно від задач, які вирішують системи управління, виділяють такі типи систем: системи стабілізації, системи програмного управління, системи слідування, адаптивні системи тощо.

З аналізу трьох видів систем (системи ЦОС, системи зв'язку, системи управління) можна зробити висновок, що скрізь в них використовуються сигнали в тому чи іншому вигляді. В різних ланках та на різних етапах роботи таких систем відбувається генерація, передача, прийом, фільтрація, оцифровка, перетворення та аналіз сигналів. Отже, без знання того, якими є принципи роботи з сигналами, неможливо будувати та використовувати ніякі електронні системи.

2.2. Лінійні дискретні системи

Всі операції обробки сигналів виконуються за допомогою систем. В цьому посібнику основна увага приділяється дискретним сигналам, а отже, далі розглянемо дискретні системи.

Дискретна система – це набір схемних компонентів або програмних процедур, які реалізують математичні операції над дискретними сигналами (рис. 2.10).

На вхід дискретної системи подаються відліки дискретного вхідного сигналу $x[n]$, а на виході формуються відліки дискретного вихідного сигналу $y[n]$.

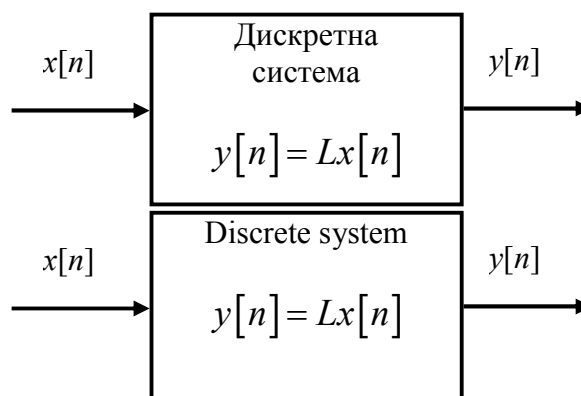


Рисунок 2.10 – Дискретна система T

З точки зору теорії сигналів, дискретна система – це *оператор або перетворення*, яке переводить вхідну послідовність відліків $x[n]$ (вхідний сигнал, вплив на систему) у вихідну послідовність відліків $y[n]$ (вихідний сигнал, реакція системи на вплив):

$$y[n] = Lx[n].$$

Попереднє співвідношення, яке позначено оператором L – це правило, або формула, або алгоритм, за яким обчислюється реакція системи на основі відліків вхідного сигналу.

Характеристики дискретних систем можна розділити на дві групи залежно від того, чи вони стосуються властивостей системи в часі, чи властивостей по відношенню до перетворень значень сигналів:

«Амплітудні» властивості:

- лінійність;
- стійкість;
- існування інверсної системи.

«Часові» властивості:

- казуальність;
- наявність пам'яті;
- стаціонарність (інваріантність у часі).

«Амплітудні» властивості дискретних систем

Система називається *лінійною*, якщо вона має дві властивості:

1. Адитивність.
2. Однорідність.

Нехай маємо систему L , і маємо сигнали $y_1[n]$ та $y_2[n]$ – реакції системи L на вхідні сигнали $x_1[n]$ та $x_2[n]$:

$$y_1[n] = Lx_1[n],$$

$$y_2[n] = Lx_2[n],$$

Властивість *адитивності* полягає в тому, що реакція системи на суму вхідних сигналів повинна дорівнювати сумі реакцій цієї системи на кожний сигнал окремо:

$$L(x_1[n] + x_2[n]) = Lx_1[n] + Lx_2[n] = y_1[n] + y_2[n].$$

Властивість *однорідності* полягає в тому, що реакція системи на масштабований вхідний сигнал (підсилений або ослаблений) дорівнює масштабованій реакції системи на немасштабований вхідний сигнал:

$$L(ax_1[n]) = aLx_1[n] = ay_1[n],$$

де a – довільна константа.

Властивості однорідності та адитивності можна записати разом в одній формулі у вигляді *принципу суперпозиції*

$$L(ax_1[n] + bx_2[n]) = aLx_1[n] + bLx_2[n] = ay_1[n] + by_2[n].$$

Принцип суперпозиції для дискретної системи T можна записати для суми довільної кількості сигналів:

$$x[n] = \sum_k a_k x_k[n], \quad y[n] = Lx[n] = \sum_k a_k y_k[n]$$

Отже, лінійна система – це система, яка має властивості адитивності та однорідності, або, що те саме, підкоряється принципу суперпозиції. Лінійність систем перетворення сигналів – фундаментальна властивість, яку має переважна більшість систем, з якими дослідники зустрічаються в практичній діяльності. При аналізі сигналів при проходженні їх через лінійні системи, внаслідок властивості адитивності, можна досліджувати реакції системи не на весь вхідний сигнал, а на його складові. Якщо ми будемо знати, як лінійна система реагує на кожен складову окремо, то ми, не проводячи вимірювань, можемо знайти, яким буде вихідний сигнал системи. Для цього треба буде тільки знайти розклад вхідного сигналу системи. Ці питання будуть більш докладно розглянуті в подальших розділах.

Система є *стійкою*, якщо її реакція на будь-який обмежений за амплітудою вхідний сигнал є обмеженою.

$$\text{якщо } \exists M_x, |x[n]| \leq M_x < +\infty \forall n, \text{ то існує } M_y, |y[n]| \leq M_y < +\infty.$$

$$\text{if } \exists M_x, |x[n]| \leq M_x < +\infty \forall n, \text{ then } M_y, |y[n]| \leq M_y < +\infty$$

Наприклад, суматор – нестійка система, в чому можна пересвідчитись, подавши на його вхід одиничну послідовність.

Система називається *інверсною*, якщо кожному вхідному сигналу відповідає єдиний вихідний сигнал. В такому випадку існує обернена система, яка може відновити вхідний сигнал за вихідним: $x[n] = L^{-1}y[n]$.

Часові властивості дискретних систем

Система називається *каузальною (детермінованою)*, якщо кожний відлік вихідного сигналу з номером n_0 залежить лише від тих відліків вхідного сигналу, номер яких не перевищує n_0 .

$$y[n] = F(x[n], x[n-1], x[n-2], \dots).$$

Система без запам'ятовування – дискретна система, кожний n -ий відлік $y[n]$ реакції якої при кожному n залежить лише від одного відліку з тим самим номером n вхідного сигналу $x[n]$:

$$y[n] = F(x[n]).$$

Якщо система має пам'ять, то вона називається динамічною системою, в іншому випадку – статичною.

Стаціонарна дискретна система – це система, для якої будь-який зсув у часі вхідного сигналу веде до такого самого зсуву вихідного сигналу. Тобто, якщо реакція системи на вхідний сигнал $x[n]$ дорівнює $y[n]$:

$$Lx[n] = y[n],$$

то реакція цієї ж системи на той самий сигнал, але який затриманий у часі на певну кількість відліків, буде рівною тій самій реакції, але яка затримана у часі на ту ж саму кількість відліків:

$$Lx[n - n_0] = y[n - n_0].$$

Стаціонарність – також дуже корисна і важлива властивість систем, і бажано, щоб всі системи, з якими має справу дослідник, були стаціонарними. Стаціонарність передбачає стабільність, незмінність властивостей системи в часі. Якщо ми подамо на вхід стаціонарної системи вхідний сигнал і отримаємо реакцію, а потім через деякий час подамо такий самий сигнал – ми отримаємо таку саму реакцію. Оскільки система стаціонарна, то одного разу визначивши її характеристики, ними можна користуватися в подальшому, оскільки вони не будуть мінятися. Стаціонарні системи називають також *інваріантними* в часі.

Наприклад:

Лінійна стаціонарна система:

$$y[n] = 3x[n] + 2.$$

Лінійна, але нестационарна, система – компресор, тобто дискретна система, яка відкидає $K-1$ з кожних K відліків вхідного сигналу, передаючи на вихід лише K -ий:

$$y[n] = x[Kn].$$

Нелінійна стаціонарна система – логарифматор:

$$y[n] = \log_{10} |x[n]|.$$

Приклади систем.

1. Ідеальна система затримки.

Ця система працює за правилом:

$$y[n] = x[n - n_d],$$

де n_d – фіксоване натуральне число, яке називається затримкою системи. Ідеальна система затримки зсуває вхідну послідовність $x[n]$ на n_d відліків.

2. Система осереднення зі зсувом (скользящего среднего, moving average).

Описується рівнянням:

$$\begin{aligned} y[n] &= \frac{1}{N_1 + N_2 + 1} \sum_{k=-N_1}^{N_2} x[n - k] = \\ &= \frac{1}{N_1 + N_2 + 1} (x[n + N_1] + x[n + N_1 - 1] + \dots + x[n] + \dots + x[n - N_2]) \end{aligned}$$

Вона розраховує n -ий відлік вихідної послідовності як середнє арифметичне $(N_1 + N_2 + 1)$ відліків вхідної послідовності, які розташовані навколо n -ого. Ця система – з запам'ятовуванням, оскільки для розрахунку поточного відліку вихідного сигналу їй потрібні не тільки поточний відлік, а ще кілька відліків вхідного сигналу.

3. Система, що розраховує квадрат сигналу.

Вона описується таким правилом:

$$y[n] = (x[n])^2.$$

Це приклад системи без запам'ятовування – вихідний відлік розраховується тільки з використанням одного відліку вхідного сигналу.

4. Суматор

$$y[n] = \sum_{k=-\infty}^n x[k].$$

Значення реакції такої системи в момент часу n дорівнює сумі всіх попередніх відліків вхідного сигналу аж до n -ого включно.

2.3. Імпульсна характеристика лінійних дискретних систем

Нехай маємо деяку лінійну систему L . Подамо на її вхід одиничний імпульс $\delta[n]$, і отримаємо реакцію $h[n]$:

$$L\delta[n] = h[n]. \quad (2.1)$$

Розглянемо деякий довільний вхідний сигнал $x[n]$. Як ми вже знаємо, довільний сигнал можна представити у вигляді лінійної комбінації (суми) зсунутих одиничних імпульсів, а саме:

$$x[n] = \sum_{k=-\infty}^{+\infty} x[k]\delta[n-k]. \quad (2.2)$$

Цей запис можна інтерпретувати як представлення нашого вхідного сигналу у вигляді суми послідовностей $\delta[n-k]$, кожна з яких підсилена в $x[k]$ разів.

Подамо сигнал $x[n]$ (2.2) на вхід системи T . Якщо ми подаємо такий сигнал на вхід лінійної системи, то реакцію системи на нього можна визначити за принципом суперпозиції:

$$y[n] = Lx[n] = L\left(\sum_{k=-\infty}^{+\infty} x[k]\delta[n-k]\right). \quad (2.3)$$

Оскільки система лінійна, для неї виконується властивість адитивності (реакція на суму вхідних сигналів буде дорівнювати сумі реакцій на кожний сигнал окремо), і можна записати:

$$y[n] = Lx[n] = L\left(\sum_{k=-\infty}^{+\infty} x[k]\delta[n-k]\right) = \sum_{k=-\infty}^{+\infty} L(x[k]\delta[n-k]).$$

Лінійність передбачає також наявність у системи властивості однорідності. Ми маємо в попередньому виразі суму реакцій системи на послідовності $\delta[n-k]$, кожна з яких помножена на числа $x[k]$ (на значення k -ого відліку вхідного сигналу). Тому рівняння (2.3) можна остаточно записати:

$$\begin{aligned} y[n] &= Lx[n] = L\left(\sum_{k=-\infty}^{\infty} x[k]\delta[n-k]\right) = \\ &= \sum_{k=-\infty}^{\infty} L(x[k]\delta[n-k]) = \\ &= \sum_{k=-\infty}^{\infty} x[k] \cdot L(\delta[n-k]). \end{aligned} \quad (2.4)$$

Отже, видно, що реакція системи на вхідний сигнал дорівнює сумі реакцій на зсунуті одиничні імпульси. Якщо система лише лінійна, то реакція на зсунуті $\delta[n]$, тобто $L(\delta[n-k])$, може залежати і від номеру відліку n , і від значення зсуву одиничних імпульсів k (тобто реакція системи на одиничні імпульси може бути різною в різні моменти часу).

Але розглянемо випадок, коли система L є не тільки лінійною, а до того ж ще і стаціонарною. В такому разі відомо, що в разі, коли $h[n]$ – реакція системи L на $\delta[n]$ (2.1), то її реакція на затриманий одиничний імпульсний сигнал $\delta[n-k]$ буде дорівнювати $h[n-k]$:

$$L\delta[n-k] = h[n-k].$$

Отже, ми можемо остаточно записати (2.4) у вигляді:

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k]. \quad (2.5)$$

Рівняння (2.5) пов'язує вихідний сигнал $y[n]$ лінійної стаціонарної дискретної системи з її вхідним сигналом $x[n]$ з допомогою реакції $h[n]$ цієї системи на одиничний імпульс.

Послідовність $h[n]$ називається імпульсною характеристикою дискретної лінійної стаціонарної системи. В подальших розділах це визначення буде доповнене умовою, що $h[n]$ буде імпульсною характеристикою лише в тому випадку, коли система була у стані спокою перед подачею на вхід одиничного імпульсу (коли початкові умови нульові). Отже, *імпульсна характеристика* – це реакція системи в стані спокою на одиничний імпульс.

З виразу (2.5) видно, що лінійна стаціонарна система повністю визначається своєю імпульсною характеристикою, тобто, маючи заздалегідь відому послідовність $h[n]$, можна обчислити реакцію $y[n]$ на будь-який вхідний сигнал $x[n]$. Для того, щоб знати, як заданий вхідний сигнал проходить через лінійну стаціонарну дискретну систему, і як вона його перетворює, не обов'язково знати структуру цієї системи, характеристики складових блоків або що. Для них не потрібне знання точного правила перетворення сигналу в системі, або оператора T . Достатньо лише для даної системи експериментально визначити або розрахувати імпульсну характеристику $h[n]$ та скористатися виразом (2.5) для отримання вихідного сигналу. Отже, особливе значення має клас дискретних систем, які одночасно є лінійними та стаціонарними, оскільки тільки для таких систем поняття імпульсної характеристики має сенс та може бути використане.

Представлення вхідного сигналу у вигляді зваженої суми зсунутих одиничних імпульсів (2.2) дозволяє інтерпретувати вихідний сигнал лінійної стаціонарної дискретної системи у вигляді суми зсунутих імпульсних характеристик (2.5), кожна з яких має коефіцієнт $x[k]$. Можна уявити, що кожний відлік вхідного сигналу з

номером k ($x[k]$), представлений як $x[k]\delta[n-k]$, в кожний k -ий момент часу викликає на виході системи реакцію у вигляді імпульсної характеристики, яка множиться на число $x[k]$. Тобто, поява на вході системи відліку $x[k]$ приводить до появи на виході системи вихідної *послідовності* $x[k]h[n-k]$. Після приходу кожного наступного вхідного відліку, на виході з'являється ще одна послідовність (імпульсна характеристика $h[n]$, підсилена в $x[k]$ разів). Для кожного k -ого вхідного відліку ці вихідні послідовності додаються системою для формування результуючого вихідного сигналу $y[n]$.

Маючи імпульсну характеристику, можна дати ще одні визначення деяких властивостей систем.

Лінійна стаціонарна дискретна система є *стійкою* лише в тому випадку, коли її імпульсна характеристика є обмеженим сигналом, тобто таким, що задовольняє умову:

$$S = \sum_{k=-\infty}^{\infty} |h[k]| < \infty.$$

Лінійна стаціонарна дискретна система є *детермінованою* лише тоді, коли її імпульсна характеристика задовольняє умову

$$h[n] = 0, \quad n < 0.$$

Надалі ми будемо працювати лише зі стійкими детермінованими системами.

Системи, імпульсна характеристика яких має лише скінченне число ненульових відліків, називається системою зі *кінцевою імпульсною характеристикою* (КІХ-система).

Системи, імпульсна характеристика яких не обмежена за тривалістю, називаються системами з *безкінченною імпульсною характеристикою* (БІХ-система).

2.4. Лінійна та кругова згортки

Якщо відліки сигналу $y[n]$ залежать від відліків послідовностей $x[n]$ та $z[n]$ за правилом (2.5), то послідовність $y[n]$ називають *дискретною згорткою* послідовностей $x[n]$ та $z[n]$ і позначають

$$y[n] = x[n] * z[n].$$

Розглянемо два сигнали $x[n]$ та $z[n]$, які є періодичними з періодом N . Сигнал-згортка цих двох послідовностей

$$y[n] = \sum_{m=0}^{N-1} x[m]z[n-m] = \sum_{m=0}^{N-1} x[n-m]z[m] \quad (2.6)$$

також буде періодичним з періодом N і називається *круговою (періодичною) згорткою* сигналів $x[n]$ та $z[n]$. Отже, достатньо розрахувати значення $y[n]$ лише для N точок, а всі інші значення будуть повторюватися.

Як видно з (2.5), всі вихідні сигнали лінійних стаціонарних дискретних систем розраховуються як результат періодичної згортки вхідної послідовності та імпульсної характеристики.

Для скінченних неперіодичних сигналів вводять поняття *лінійної (аперіодичної) згортки*. Нехай $x[n]$ та $h[n]$ – скінченні сигнали тривалістю відповідно N_1 та N_2 відліків. Лінійною (аперіодичною) згорткою цих сигналів буде дискретний сигнал $y[n]$, який визначається:

$$y[n] = \sum_{m=0}^n x[m]h[n-m] = \sum_{m=0}^n x[n-m]h[m]. \quad (2.7)$$

Вираз (2.7) можна записати зі сталими а не змінними межами додавання, але тоді в ньому будуть присутні нульові доданки:

$$y[n] = \sum_{m=0}^{N_1+N_2-1} x[m]h[n-m] = \sum_{m=0}^{N_1+N_2-1} x[n-m]h[m].$$

В цьому випадку сигнал $z[n]$ буде мати довжину $N' = N_1 + N_2 - 1$, тобто буде на один відлік коротше, ніж сума тривалостей двох сигналів, для яких розраховується згортка.

Як правило, в комп'ютерних системах обчислення згорток для сигналів великої тривалості проводиться не безпосередньо з використанням цих виразів, а з використанням перетворення Фур'є, оскільки для його обчислення існують швидкі алгоритми. В дискретних системах, особливо в тих, які працюють в реальному часі, зазвичай користуються тим, що операція згортки двох сигналів в часовій області еквівалентна множенню їх спектрів в частотній області. Тому для обчислення згортки спочатку розраховують спектри двох сигналів, потім перемножують їх в частотній області, а після цього – виконують обернене перетворення Фур'є із спектральної області в часову. Результат цього оберненого перетворення і буде рівний згортці двох початкових сигналів.

Алгоритми для обчислення кругової згортки можна застосувати і для обчислення лінійної згортки. Для цього необхідно доповнити нулями обидві послідовності так, щоб вони мали однакову довжину.

Властивості згортки.

Комутативність

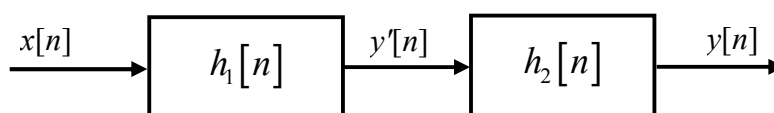
$$x[n] * h[n] = h[n] * x[n].$$

Згортка є *дистрибутивною* відносно додавання:

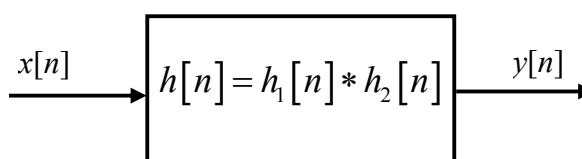
$$x[n] * (h_1[n] + h_2[n]) = x[n] * h_1[n] + x[n] * h_2[n].$$

При каскадному (послідовному) з'єднанні (рис. 2.11, а) дві лінійні стаціонарні дискретні системи з імпульсними характеристиками $h_1[n]$ та $h_2[n]$ можна замінити еквівалентною лінійною стаціонарною дискретною системою, імпульсна характеристика якої буде визначатися як згортка імпульсних характеристик двох систем (рис. 2.11, б):

$$h[n] = h_1[n] * h_2[n].$$



а)

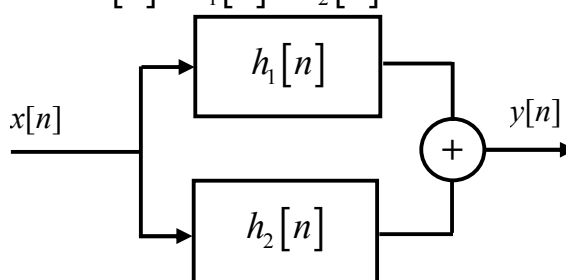


б)

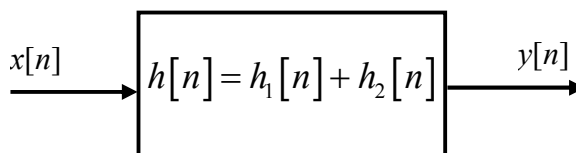
Рисунок 2.11 – Послідовне з'єднання систем

При паралельному з'єднанні (рис. 2.12, а) дві лінійні стаціонарні дискретні системи з імпульсними характеристиками $h_1[n]$ та $h_2[n]$ мають спільний вхід, а їх вихідні послідовності додаються. Їх можна замінити еквівалентною лінійною стаціонарною дискретною системою, імпульсна характеристика якої дорівнює сумі імпульсних характеристик (рис. 2.12, б):

$$h[n] = h_1[n] + h_2[n].$$



а)



б)

Рисунок 2.12 – Паралельне з'єднання систем

2.5. Опис лінійних дискретних систем за допомогою різницевих рівнянь

Як відомо з теорії електричних кіл, зв'язок між вхідними та вихідними сигналами лінійної неперервної системи може бути виражений у вигляді диференціального рівняння зі сталими коефіцієнтами, а отже, аналогова електрична схема описується наборами коефіцієнтів цього рівняння a_i та b_i .

Аналогічно, існує важливий підклас лінійних дискретних систем (стаціонарні системи з раціональною характеристичною функцією), для яких пара «вхідний сигнал – вихідний сигнал» пов'язана лінійним різницеvim рівнянням порядку N з постійними коефіцієнтами:

$$\sum_{k=0}^N a_k y[n-k] = \sum_{m=0}^M b_m x[n-m], \quad (2.8)$$

де a_k та b_m – коефіцієнти рівняння;

$x[n]$ – вхідний сигнал;

$y[n]$ – вихідний сигнал.

Структура різницевого рівняння схожа на структуру диференційного рівняння аналогової лінійної системи, але замість операції диференціювання в формулі (2.8) фігурують затримки дискретних послідовностей на k або m відліків. На відміну від неперервного випадку, в дискретній системі відсутні будь-які обмеження на співвідношення між N та M – кількостями вхідних та вихідних відліків, які використовуються при обчисленнях вихідного сигналу.

Якщо відоме різницеве рівняння, яке описує дану систему, то можемо розрахувати значення вихідного сигналу по заданому вхідному сигналу:

$$\begin{aligned} a_0 y[n] + \sum_{k=1}^N a_k y[n-k] &= \sum_{m=0}^M b_m x[n-m], \\ y[n] &= \frac{1}{a_0} \left(\sum_{m=0}^M b_m x[n-m] - \sum_{k=1}^N a_k y[n-k] \right). \end{aligned} \quad (2.9)$$

Рівняння «відоме» – це значить, що відомі всі коефіцієнти рівняння a_k та b_m , тобто числа, які потрібні для проведення розрахунків за цим рівнянням.

Якщо розглянути праву частину (2.9), то видно, що для розрахунку значення поточного n -ого відліку вихідного сигналу $y[n]$, нам потрібні M відліків вхідного сигналу $x[n]$ (включаючи поточний n -ий), а також N попередні відліки вихідного сигналу.

Розглянемо, як проводити розрахунки відліків вихідного сигналу за допомогою різницевого рівняння. В початковий момент часу, коли на вхід системи потрапляє тільки нульовий відлік вхідного сигналу $x[0]$, і ми маємо розрахувати нульовий відлік вихідного сигналу $y[0]$. Запишемо його значення через рівняння (2.9):

$$\begin{aligned} y[0] &= \frac{1}{a_0} \left(\sum_{m=0}^M b_m x[0-m] - \sum_{k=1}^N a_k y[0-k] \right) = \\ &= \frac{1}{a_0} (b_0 x[0] + b_1 x[-1] + \dots + b_{M-1} x[-M] - \\ &\quad - a_1 y[-1] - a_2 y[-2] - \dots - a_{N-1} y[-N]). \end{aligned}$$

Видно, що для того, щоб розрахувати значення, треба задати значення всіх коефіцієнтів різницевого рівняння, а також всі значення попередніх відліків вхідного та вихідного сигналу (з від'ємними номерами). Щодо коефіцієнтів, то якщо ми знаємо, з якою системою працюємо, то коефіцієнти різницевого рівняння нам відомі. Щодо попередніх відліків вхідного сигналу, то, оскільки ми розглядаємо початковий момент часу, то всі попередні відліки вхідного сигналу дорівнюють нулю, оскільки сигнал тільки почав надходити на вхід нашої системи:

$$x[n] \equiv 0 \text{ для всіх } n \leq 0.$$

Але для розрахунків ми не знаємо значень попередніх відліків вихідного сигналу $y[n]$, які також входять у рівняння. Ці відліки – попередні відліки вихідного сигналу системи, які були (або могли бути) на виході до того, як на вхід поступив наш вхідний сигнал. Ці відліки описують стан системи до початку надходження вхідного сигналу $x[n]$, і цей стан нам невідомий, тому у визначенні відліків вихідного сигналу $y[n]$ є неоднозначність:

$$\begin{aligned} y[0] &= \frac{1}{a_0} \left(\underbrace{b_0 x[0]}_{\text{відоме}} - \underbrace{a_1 y[-1] - a_2 y[-2] - \dots - a_{N-1} y[-N]}_{\text{невідомі}} \right) \\ y[0] &= \frac{1}{a_0} \left(\underbrace{b_0 x[0]}_{\text{known}} - \underbrace{a_1 y[-1] - a_2 y[-2] - \dots - a_{N-1} y[-N]}_{\text{unknown}} \right) \end{aligned}$$

В цьому виразі відомо, чому дорівнює $x[0]$, всі коефіцієнти a_k та b_m , але маємо також N невідомих значень попередніх відліків вихідної послідовності $y[-1], y[-2], \dots, y[-N]$, які потрібні для однозначного знаходження відліків $y[n]$.

Для того, щоб зняти цю невизначеність у розрахунках, треба задати N додаткових умов до рівняння (2.9). Часто ці додаткові умови бувають задані у вигляді попередніх значень деякого умовного вихідного сигналу $y[-1], y[-2], \dots, y[-N]$, які були на виході досліджуваної системи до того, як на вхід почали поступати значення нашого вхідного сигналу $x[n]$. Тобто, може бути заданий стан системи до того, як на вхід було подано вхідний сигнал. В цьому випадку можна розрахувати всі відліки вихідної послідовності $y[n]$ за рекурентною формулою (2.9) за відомими відліками вхідного сигналу $x[n]$.

Якщо система, що описується лінійним різницеvim рівнянням, не тільки лінійна, а ще стаціонарна та детермінована, то розв'язання рівняння за наявності додаткових умов буде однозначним. Тобто, одне рівняння може відповідати лише одній такій системі.

Якщо система **детермінована**, тобто поточний вихідний відлік повинен залежати лише від поточного вхідного відліку та попередніх вхідних і вихідних, то з умови $x[n] = 0$ для всіх $n < 0$ витікає рівність $y[n] = 0$ для $n < 0$.

В цьому випадку говорять про *нульові початкові умови* (стан спокою системи):

$$y[-1] = 0, y[-2] = 0, \dots, y[-N] = 0.$$

Вихідний сигнал лінійної стаціонарної детермінованої дискретної системи з нульовими початковими умовами обчислюються однозначно за різницеvim рівнянням, наприклад:

$$\begin{aligned} n = 0: \quad y[0] &= \frac{1}{a_0} \left(\sum_{m=0}^M b_m x[n-m] - \sum_{k=1}^N a_k y[n-k] \right) = \frac{b_0}{a_0} x[0], \\ n = 1: \quad y[1] &= \frac{1}{a_0} (b_0 x[1-0] + b_1 x[1-1] - a_1 y[1-1]) = \frac{b_0}{a_0} x[1] + \frac{b_1}{a_0} x[0] - \frac{a_1}{a_0} y[0], \\ n = 2: \quad y[2] &= \frac{1}{a_0} (b_0 x[2-0] + b_1 x[2-1] + b_2 x[2-2] - a_1 y[2-1] - a_2 y[2-2]) = \\ &= \frac{b_0}{a_0} x[2] + \frac{b_1}{a_0} x[1] + \frac{b_2}{a_0} x[0] - \frac{a_1}{a_0} y[1] - \frac{a_2}{a_0} y[0], \\ n = 3: \quad y[3] &= \frac{1}{a_0} (b_0 x[3-0] + b_1 x[3-1] + b_2 x[3-2] + b_3 x[3-3] - \\ &\quad - a_1 y[3-1] - a_2 y[3-2] - a_3 y[3-3]) = \\ &= \frac{b_0}{a_0} x[3] + \frac{b_1}{a_0} x[2] + \frac{b_2}{a_0} x[1] + \frac{b_3}{a_0} x[0] - \frac{a_1}{a_0} y[2] - \frac{a_2}{a_0} y[1] - \frac{a_3}{a_0} y[0], \\ &\dots \end{aligned}$$

За такою процедурою вихідний сигнал системи розраховується рекурсивно, тобто кожен наступний відлік вихідного сигналу виражається через попередні та поточний відліки вхідного сигналу та визначені раніше попередні відліки вихідного сигналу. Отже, в загальному випадку вихідний сигнал лінійної дискретної системи визначається по вхідному неоднозначно, і для проведення розрахунків потрібна додаткова інформація. Якщо така інформація задана у вигляді початкових умов, то розрахунки відліків вихідного сигналу можна проводити за рекурсивним рівнянням. В подальшому ми будемо працювати лише з системами, для яких початкові умови – нульові.

Тепер можна пояснити уточнення, яке було зроблено раніше (див. розділ 2.3), коли вводилося визначення імпульсної характеристики. Якщо ми подамо на вхід системи одиничний імпульс, то, в залежності від початкових умов, ми можемо тримати різні реакції на неї. І лише у випадку, коли початкові умови були нульовими (система знаходилась в стані спокою), реакція на одиничний імпульс буде саме імпульсною характеристикою.

2.6. Z- перетворення

Властивості дискретних систем доцільно та зручно вивчати з допомогою z-перетворення. z-перетворення дискретних сигналів аналогічне до перетворення Лапласа неперервних сигналів. Ці перетворення мають схожий зв'язок з перетворенням Фур'є.

z-перетворення сигналу $x[n]$ – це правило, яке ставить у відповідність сигналу $x[n]$ деяку функцію $X(z)$ від неперервної комплексної змінної z . Часто $X(z)$ позначають $Z(x[n])$:

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n}. \quad (2.10)$$

Права частина формули – нескінченний степеневий ряд від комплексної змінної z . z-перетворення ставить у відповідність кожному дискретному сигналу в часовій області деяку неперервну комплекснозначну функцію в z -області. В загальному випадку, в разі, коли початковий сигнал скінченний та обмежений (як всі реальні дискретні сигнали), то його z -перетворення збіжне. Обернене z -перетворення ґрунтується на інтегральній теоремі Коші.

У випадку, коли сигнал скінченний, його z -перетворення можна записати:

$$X(z) = \sum_{n=0}^{N-1} x[n]z^{-n}. \quad (2.11)$$

z-перетворення є найбільш корисним, коли нескінченна сума (2.10) може бути записана в компактному вигляді, особливо якщо це можна зробити у вигляді раціональної функції, тобто

$$X(z) = \frac{P(z)}{Q(z)}.$$

Відповідність між сигналом та його z-перетворенням позначається $x[n] \xleftrightarrow{Z} X(z)$.

Нехай маємо два дискретних сигнали з їх z-перетвореннями:

$$x_1[n] \xleftrightarrow{Z} X_1(z), \quad x_2[n] \xleftrightarrow{Z} X_2(z).$$

Розглянемо основні властивості z-перетворення:

Лінійність

$$ax_1[n] + bx_2[n] \xleftrightarrow{Z} aX_1(z) + bX_2(z).$$

z-перетворення затриманого сигналу

$$x[n - n_0] \xleftrightarrow{Z} z^{-n_0} X(z)$$

Доведемо властивість затримки: нехай $y[n] = x[n - n_0]$, тоді

$$\begin{aligned} Y(z) &= \sum_{n=-\infty}^{\infty} x[n - n_0] z^{-n} = \sum_{m=n-n_0}^{\infty} x[m] z^{-(m+n_0)} = \\ &= z^{-n_0} \sum_{m=-\infty}^{\infty} x[m] z^{-m} = z^{-n_0} X(z). \end{aligned}$$

z-перетворення згортки

Нехай маємо сигнал, який є згорткою двох сигналів:

$$y[n] = x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k] h[n - k].$$

Обчислимо його z-перетворення. За визначенням (2.10) маємо:

$$Y(z) = \sum_{n=-\infty}^{\infty} y[n] z^{-n} = \sum_{n=-\infty}^{\infty} \left(\sum_{k=-\infty}^{\infty} x[k] h[n - k] \right) z^{-n}.$$

Оскільки змінні n та k незалежні, то поміняємо порядок додавання

$$Y(z) = \sum_{k=-\infty}^{\infty} \left(\sum_{n=-\infty}^{\infty} x[k] h[n - k] \right) z^{-n},$$

а також винесемо множник $x[k]$ з внутрішньої суми, оскільки він не залежить від індексу n і буде константою для кожного n :

$$Y(z) = \sum_{k=-\infty}^{\infty} x[k] \sum_{n=-\infty}^{\infty} h[n - k] z^{-n}.$$

Зробимо у внутрішній сумі заміну змінних $m = n - k$, і тоді матимемо:

$$Y(z) = \sum_{k=-\infty}^{+\infty} x[k] \left(\sum_{m=-\infty}^{+\infty} h[m] z^{-(m+k)} \right) = \sum_{k=-\infty}^{+\infty} x[k] z^{-k} \left(\sum_{m=-\infty}^{+\infty} h[m] z^{-m} \right) = X(z) H(z).$$

Отже, остаточно можна записати:

$$x[n] * h[n] \xleftrightarrow{Z} X(z) \cdot H(z), \quad (2.12)$$

тобто, згортка двох сигналів в часовій області відповідає добутку їх z-перетворень.

Множення на експоненціальну послідовність

$$z_0^n x[n] \xleftrightarrow{z} X\left(\frac{z}{z_0}\right).$$

Диференціювання $X(z)$

$$nx[n] \xleftrightarrow{z} -z \frac{dX(z)}{dz}.$$

z- перетворення спряженої послідовності

$$x^*[n] \xleftrightarrow{z} X^*(z^*).$$

Обернення послідовності в часі

$$x^*[-n] \xleftrightarrow{z} X^*\left(\frac{1}{z^*}\right).$$

z-перетворення є основою для перетворення лінійних різницевих рівнянь з постійними коефіцієнтами в алгебраїчні рівняння.

2.7. Характеристична функція лінійних дискретних систем

Згадаємо, як пов'язані вихідний сигнал лінійної стаціонарної дискретної системи з її імпульсною характеристикою:

$$y[n] = x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k] h[n-k], \quad (2.13)$$

тобто вихідний сигнал є результатом згортки вхідного сигналу та імпульсної характеристики системи. Застосуємо z-перетворення до лівої та правої частин рівняння (2.13) і згадаємо властивості z-перетворення затриманої послідовності $h[n-k]$:

$$Y(z) = X(z) \cdot Z(h[n-k]) = X(z) \cdot \sum_{k=-\infty}^{+\infty} h[k] z^{-k} = X(z) \cdot H(z),$$

звідки

$$H(z) = \frac{Y(z)}{X(z)}, \quad (2.14)$$

де $H(z)$ – z-перетворення імпульсної характеристики системи.

Характеристична функція $H(z)$ – це відношення z-перетворень вихідного та вхідного сигналів. Вона дорівнює z-перетворенню імпульсної характеристики системи.

2.8. Зв'язок між характеристичною функцією та різницеvim рівнянням

Оскільки система може бути задана не тільки своєю імпульсною характеристикою, а і різницеvim рівнянням, то знайдемо зв'язок між цими двома способами описання одної і тієї ж самої системи. Застосуємо z-перетворення до правої та лівої частини різницевого рівняння (2.8) та використаємо властивість z-перетворення затриманої послідовності:

$$\begin{aligned}
 Z\left(\sum_{k=0}^{N-1} a_k y[n-k]\right) &= Z\left(\sum_{m=0}^{M-1} b_m x[n-m]\right) = \\
 &= \sum_{k=0}^{N-1} Z(a_k y[n-k]) = \sum_{m=0}^{M-1} Z(b_m x[n-m]) = \\
 &= \sum_{k=0}^{N-1} a_k (Z(y[n-k])) = \sum_{m=0}^{M-1} b_m (Z(x[n-m])) = \\
 &= \sum_{k=0}^{N-1} a_k z^{-k} Y(z) = \sum_{m=0}^{M-1} b_m z^{-m} X(z).
 \end{aligned}$$

Оскільки $Y(z)$ та $X(z)$ не залежать від індексів k та m , їх можна винести за знак суми:

$$Y(z) \left(\sum_{k=0}^{N-1} a_k z^{-k} \right) = X(z) \left(\sum_{m=0}^{M-1} b_m z^{-m} \right).$$

Поділимо ліву частину останнього рівняння на праву, щоб отримати відношення z-перетворень вихідного та вхідного сигналів, і остаточно отримаємо:

$$\frac{Y(z)}{X(z)} = \frac{\sum_{m=0}^{M-1} b_m z^{-m}}{\sum_{k=0}^{N-1} a_k z^{-k}}. \quad (2.15)$$

Якщо порівняти вирази (2.15) та (2.14), то можна отримати зв'язок характеристичної функції $H(z)$ з коефіцієнтами різницевого рівняння a_k та b_m :

$$H(z) = \frac{\sum_{m=0}^{M-1} b_m z^{-m}}{\sum_{k=0}^{N-1} a_k z^{-k}}. \quad (2.16)$$

З (2.16) видно, що різницеве рівняння можна однозначно відновити за характеристичною функцією, яка представлена раціональним дробом з поліномами відносно змінної z в чисельнику та знаменнику. І навпаки, можна побудувати характеристичну функцію, маючи коефіцієнти різницєвого рівняння.

2.9. Комплексна частотна характеристика лінійних дискретних систем

Нехай маємо лінійну стаціонарну систему з імпульсною характеристикою $h[n]$. Подамо на її вхід сигнал $x[n] = e^{j\omega n}$, $-\infty < n < \infty$ – комплексний експоненціальний сигнал одиничної амплітуди з частотою ω рад/с. Згідно властивості (2.7), вихідний сигнал буде розраховуватися через згортку та мати вигляд:

$$y[n] = \sum_{k=-\infty}^{+\infty} e^{j\omega k} h[n-k] = \sum_{k=-\infty}^{+\infty} h[k] e^{j\omega[n-k]} = \sum_{k=-\infty}^{+\infty} h[k] e^{j\omega n} e^{-j\omega k} = e^{j\omega n} \sum_{k=-\infty}^{+\infty} h[k] e^{-j\omega k}.$$

Якщо покласти

$$H(j\omega) = \sum_{k=-\infty}^{+\infty} h[k] e^{-j\omega k}, \quad (2.17)$$

то отримаємо вихідний сигнал системи

$$y[n] = H(j\omega) e^{j\omega n}.$$

Видно, що вихідний сигнал системи при подачі на вхід комплексного синусоїдального сигналу частоти ω буде сигналом такої самої частоти. Амплітуда цієї синусоїди буде дорівнювати значенню комплексної функції $H(j\omega)$ на частоті ω . Функція $H(j\omega)$ може бути отримана з імпульсної характеристики $h[n]$ за (2.17) (в подальших розділах буде показано, що (2.17) – це перетворення Фур'є для нескінченної послідовності).

$H(j\omega)$ – комплексна частотна характеристика (КЧХ) лінійної дискретної стаціонарної системи. За визначенням КЧХ – це залежність (комплексного) коефіцієнту передачі системи від частоти. Вона показує, як змінюється комплексна амплітуда вхідного експоненціального сигналу при проходженні через дану систему.

Комплексна частотна характеристика системи може бути записана в алгебраїчній формі через свою дійсну та уявну частини:

$$H(j\omega) = H_R(j\omega) + jH_I(j\omega),$$

або у показниковому вигляді:

$$H(j\omega) = |H(j\omega)| e^{j\arg(H(j\omega))} = K(\omega) e^{j\Phi(\omega)},$$

де $K(\omega)$ – модуль КЧХ – амплітудно-частотна характеристика (АЧХ) – це залежність коефіцієнту передачі системи від частоти. АЧХ показує, як змінюється

амплітуда вхідного комплексного експоненціального сигналу при проходженні через систему залежно від того, яка частота цього сигналу;

$\Phi(\omega)$ – аргумент КЧХ – *фазо-частотна характеристика* (ФЧХ) – це залежність фазового зсуву між вихідним і вхідним сигналом від частоти цього сигналу. Вона показує, як змінюється фаза вхідного комплексного експоненціального сигналу при проходженні через систему залежно від того, яка частота цього сигналу.

Треба зауважити деякі термінологічні уточнення. КЧХ називають комплексним коефіцієнтом передачі системи. Залежно від того, якими є сигнали на вході, може іти мова про коефіцієнт передачі струму або напруги. В цьому випадку правильно буде говорити, що $K(\omega)$ – це амплітудно-частотна характеристика не системи, а АЧХ коефіцієнту передачі напруги або струму. І аналогічно, $\Phi(\omega)$ – є ФЧХ коефіцієнту передачі. Далі в цій книзі ми будемо вважати, що вхідні сигнали – це напруга, і тому будемо називати АЧХ і ФЧХ скорочено, маючи на увазі коефіцієнт передачі напруги.

Для різних частот вхідного синусоїдального сигналу ω КЧХ показує, як змінюється амплітуда (в $K(\omega)$ разів) і фаза (на $\Phi(\omega)$ радіан) цього сигналу при проходженні через лінійну дискретну систему. Залежно від того, яку частоту має вхідний синусоїдальний сигнал, зміна його фази та амплітуди після проходження через систему буде різною. Але частота залишиться такою самою. Якщо спектр вхідного сигналу буде

$$X(j\omega) = |X(j\omega)| e^{j\arg(X(j\omega))},$$

то можна записати вираз для спектру результуючого сигналу:

$$Y(j\omega) = H(j\omega)X(j\omega) = |H(j\omega)| e^{j\arg(H(j\omega))} \cdot |X(j\omega)| e^{j\arg(X(j\omega))} = |Y(j\omega)| e^{j\arg(Y(j\omega))}.$$

Амплітудний спектр вихідного сигналу буде рівним

$$|Y(j\omega)| = |H(j\omega)| \cdot |X(j\omega)|,$$

а фазовий спектр вихідного сигналу

$$\arg(Y(j\omega)) = \arg(H(j\omega)) + \arg(X(j\omega)).$$

З іншого боку, можна записати, що

$$\Phi(\omega) = \arg(Y(j\omega)) - \arg(X(j\omega)),$$

$$K(\omega) = \frac{|Y(j\omega)|}{|X(j\omega)|}.$$

Згадаємо, що дискретну систему можна описувати деяким оператором, тобто правилом, яке пов'язує вхідний і вихідний сигнал та показує, як сигнал перетворюється системою. З математики відомо, що власною функцією для оператора є функція, для якої результат дії на неї оператора є тією ж самою функцією, з точністю до постійного множника:

$$Lf = \lambda f,$$

де f – власна функція;

λ – власне значення.

Якщо повернутися до розгляду дискретної системи, то можна сказати, що сигнал $e^{j\omega n}$ – є власна функція системи з власним значенням $H(j\omega)$. Значення $H(j\omega)$ описує функціональну залежність зміни комплексної амплітуди вхідного сигналу від частоти ω .

Комплексні експоненціальні послідовності є власними функціями лінійних стаціонарних систем, а реакція такої системи на синусоїдальну послідовність залишається синусоїдальною послідовністю з такою самою частотою, фаза і амплітуда якої повністю визначаються системою. Через це синусоїдальні (або, що те ж саме) комплексні експоненціальні сигнали відіграють особливо важливу роль в теорії сигналів та в аналізі перетворення сигналів в лінійних системах.

Знаючи, що такі сигнали при проходженні через систему не міняють свою частоту, ми можемо звести дослідження проходження складного сигналу через цю лінійну систему до проходження окремих його синусоїдальних складових. Для цього ми можемо розкласти довільний вхідний сигнал на синусоїдальні складові різних частот, кожна з яких буде мати свою амплітуду та початкову фазу.

Якщо знати зміни, яких зазнають амплітуда та фаза кожної синусоїдальної складової сигналу при проходженні через систему (тобто амплітудну та фазову характеристики), то можна визначити, якими вони будуть на виході. Потім, склавши всі складові на виході системи, можна отримати результат проходження всього довільного вхідного сигналу.

Саме через те, що синусоїдальні сигнали є власними функціями операторів, які слугують математичними моделями лінійних систем, в теорії сигналів набуло такої популярності перетворення Фур'є, яке дозволяє представити довільний сигнал у вигляді сукупності гармонічних складових.

2.10. Еквівалентність способів описання систем

Отже, для одної і тієї самої лінійної дискретної стаціонарної системи можна отримати чотири способи описати її роботу, тобто чотири способи пов'язати вхідний та вихідний сигнал системи:

- різницеве рівняння;
- імпульсна характеристика;
- характеристична функція;
- комплексна частотна характеристика.

Всі способи є еквівалентними, оскільки вони повністю характеризують дану систему: маюче одне описання, можна отримати інші три за шляхом деяких перетворень.

Наприклад, якщо маємо різницеве рівняння, то подавши на вхід одиничний імпульс та задавши нульові початкові умови, отримаємо імпульсну характеристику. Взявши коефіцієнти цього різницевого рівняння, можна за ними записати вираз для характеристичної функції. Якщо виконати пряме перетворення Фур'є імпульсної характеристики, то отримаємо КЧХ системи.

2.11. Висновки до розділу 2 та рекомендована література

В цьому розділі було розглянуто дискретні системи та способи їх описання. Переважна більшість систем є або принаймні вважається лінійними, тобто такими, які підкоряються принципу суперпозиції. Властивості стаціонарних лінійних систем та їх реакцію на вхідні сигнали зручно описувати чотирма еквівалентними способами:

- імпульсною характеристикою;
- різницеvim рівнянням;
- характеристичною функцією;
- комплексною частотною характеристикою.

Маючи одну з вищезгаданих характеристик, можна знайти всі інші. Важливим математичним співвідношенням за яким можна знайти вихідний сигнал системи при наявних імпульсній характеристиці та вхідному сигналі, є згортка. Корисним для аналізу дискретних систем і сигналів є z -перетворення, яке є узагальненням перетворення Лапласа для випадку дискретних сигналів. z -перетворення також пов'язує між собою імпульсну характеристику та характеристичну функцію лінійної дискретної системи.

В цьому розділі отримано також вирази для комплексної частотної характеристики системи, та для її АЧХ і ФЧХ. Також було проаналізовано, якими є власні функції лінійних дискретних систем, і визначено, що це – комплексні синусоїдальні коливання. При проходженні через систему вони не змінюють свою частоту, а міняється лише амплітуда та фаза. Через це розклад сигналів на гармонічні складові та дослідження їх проходження через лінійні системи є одним з найбільш розповсюджених методів аналізу сигналів.

Про види та характеристики дискретних систем можна прочитати в [2-3]. Зокрема, про системи дослідження біомедичних сигналів докладно розповідається у [19]. Докладний розгляд теорії дискретних систем та їх характеристик подано в класичному підручнику з цифрової обробки сигналів [2]. Питання власних функцій дискретних систем розглянуто в [20].

2.12. Питання для самоперевірки

1. З яких блоків складається система цифрової обробки дискретних сигналів? Яке призначення кожного з них?
2. З яких блоків складається система управління? Яке призначення кожного з них?
3. З яких блоків складається система зв'язку? Яке призначення кожного з них?
4. Які дискретні системи називаються лінійними? В чому полягають властивості стаціонарності, детермінованості, стійкості систем?
5. Що таке імпульсна характеристика лінійних дискретних систем, як її отримати?
6. Як розраховується згортка сигналів, які види згортки бувають? Як, маючи імпульсну характеристику системи та вхідний сигнал, розрахувати відліки вихідного сигналу?
7. Які сигнали для лінійної дискретної системи пов'язує між собою різницеве рівняння? Що таке початкові умови і якими вони є для детермінованих лінійних дискретних систем?
8. Що таке z -перетворення і які його властивості?
9. Що таке характеристична функція і як вона пов'язана з різницеvim рівнянням і імпульсною характеристикою?
10. Що таке комплексна частотна характеристика дискретної системи, як її отримати? Що таке АЧХ та ФЧХ системи, що вони показують?
11. Як, маючи вхідний сигнал системи і її комплексну частотну характеристику отримати вихідний сигнал системи?
12. Що таке власні функції лінійної дискретної стаціонарної системи, яке значення вони мають для аналізу сигналів та систем?

3. СПЕКТРАЛЬНИЙ АНАЛІЗ СИГНАЛІВ ЗА ФУРЬЄ

3.1. Аналіз неперервних періодичних сигналів

Equation Chapter 1 Section 3

Як було показано в попередньому розділі, складні сигнали, які треба досліджувати, можна представляти у вигляді розкладів на елементарні більш прості складові. Таке представлення отримується як розклад функції, що описує наш сигнал за системою деяких обраних заздалегідь базисних функцій. Оскільки кожна така система має свої особливі властивості, і представлення сигналів за допомогою цих функцій будуть відрізнятися. І кожне представлення буде виявляти в сигналі ті чи інші властивості, які можуть бути корисними для дослідника. Вибір базису для аналізу сигналів повинен ґрунтуватися на тому, які саме властивості сигналу треба дослідити; знаючи це, треба обирати або створювати базисні функції для аналізу.

Найбільш розповсюдженою системою, за якою проводять розклад сигналів, є система тригонометричних функцій (синусоїд та косинусоїд) з частотами, які кратні одній основній частоті.

Розглянемо випадок, коли сигнал, який треба проаналізувати, є неперервним та періодичним сигналом довільної форми, період якого дорівнює T_0 . З цим періодом

пов'яжемо частоту ω_0 , яка дорівнює $\frac{2\pi}{T_0}$, та сформуємо набір синусоїдальних та

косинусоїдальних функцій (рис. 3.1):

$$1, \sin(\omega_0 t), \cos(\omega_0 t), \sin(2\omega_0 t), \cos(2\omega_0 t), \dots, \sin(n\omega_0 t), \cos(n\omega_0 t), \dots$$

В більш загальному вигляді таку систему функцій можна записати у вигляді комплексних експоненціальних функцій:

$$\dots e^{-j2\omega_0 t}, e^{-j\omega_0 t}, 1, e^{j\omega_0 t}, e^{j2\omega_0 t}, \dots \quad (3.1)$$

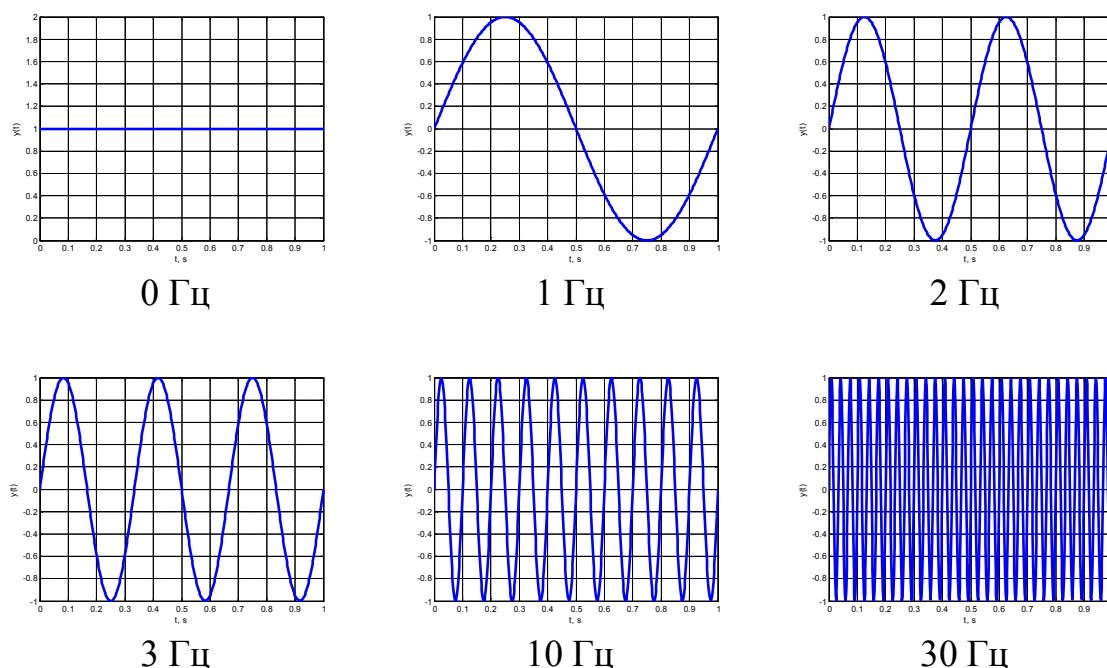


Рисунок 3.1 – Приклади базисних функцій ортогонального тригонометричного базису

За формулами Ейлера

$$e^{j\omega t} = \cos \omega t + j \sin \omega t$$

$$e^{-j\omega t} = \cos \omega t - j \sin \omega t$$

кожну таку комплексну експоненціальну функцію можна представити як суму двох синусоїдальних коливань. Такі функції є в загальному випадку нескінченними (визначеними для проміжку часу від $-\infty$ до $+\infty$), а також неперервними. Отже, якщо ми хочемо представляти за допомогою суми таких функцій деякий досліджуваний сигнал, то такий сигнал також має бути неперервним і нескінченним. Для сигналів інших видів розклад за системою (3.1) неможливий.

Можна довести, що всі функції (3.1) ортогональні на інтервалі, який співпадає з періодом основної частоти $T_0 = \frac{2\pi}{\omega_0}$. Набір функцій (3.1) називається

тригонометричним базисом Фур'є і скорочено позначається $\{e^{jn\omega_0 t}\}_{n=-\infty}^{\infty}$. Розклад сигналу за таким базисом називається *гармонічним аналізом*.

Отже, нехай маємо деякий неперервний періодичний з періодом T_0 сигнал $x(t)$ зі скінченною енергією $x(t) \in L^2(R)$. Вимоги скінченності енергії для сигналів формулюються у вигляді відомих з математики *умов Діріхле*: функція, яка описує цей сигнал, повинна бути обмеженою за амплітудою, кусково-неперервною та мати

в області визначення скінченну кількість екстремумів. Треба звернути увагу, що період сигналу, який аналізується, повинен співпадати з періодом основної частоти в базисі.

Отримаємо вираз для представлення сигналу $x(t)$ в тригонометричному базисі Фур'є (3.1). Як відомо, для будь-якого ортогонального базису $\{\phi_n(t)\}_{n=-\infty}^{+\infty}$ можна записати загальний вираз для розкладу в ньому:

$$x(t) = \sum_{n=-\infty}^{+\infty} c_n \phi_n(t) = \sum_{n=-\infty}^{+\infty} \frac{\langle x(t), \phi_n(t) \rangle}{\|\phi_n(t)\|^2} \phi_n(t).$$

Тут

$$c_n = \frac{\langle x(t), \phi_n(t) \rangle}{\|\phi_n(t)\|^2}.$$

Для нашого випадку розклад за системою функцій $\{e^{jn\omega_0 t}\}_{n=-\infty}^{\infty}$ матиме вигляд:

$$x(t) = \sum_{n=-\infty}^{+\infty} c_n e^{jn\omega_0 t} = \dots c_{-2} e^{-j2\omega_0 t} + c_{-1} e^{-j\omega_0 t} + c_0 + c_1 e^{j\omega_0 t} + c_2 e^{j2\omega_0 t} + \dots \quad (3.2)$$

Запис (3.2) є *рядом Фур'є в комплексній формі*. Знайдемо коефіцієнти розкладу c_n . Визначимо спочатку евклідову норму базисної функції для довільного номеру n . Елементи базису (3.1) та сигнал $x(t)$ є векторами в Гільбертовому просторі, а в ньому норма може бути визначена як квадратний корінь зі скалярного добутку функції самої на себе. Скалярний добуток векторів в такому просторі визначається як

$$\langle x, y \rangle = \int_T x(t) y^*(t) dt.$$

Норма базисної функції буде розрахована за виразом:

$$\|e^{jn\omega_0 t}\| = \sqrt{\langle e^{jn\omega_0 t}, e^{jn\omega_0 t} \rangle} = \sqrt{\int_{-T/2}^{T/2} e^{jn\omega_0 t} \cdot e^{-jn\omega_0 t} dt} = \sqrt{\int_{-T/2}^{T/2} e^{j0} dt} = \sqrt{T}.$$

Отже, коефіцієнти розкладу сигналу в базисі Фур'є для неперервних періодичних сигналів можна визначити:

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t) \cdot e^{-jn\omega_0 t} dt. \quad (3.3)$$

Тут інтегрування ведеться по одному періоду, оскільки функції періодичні та нескінченні. Від'ємний показник степеню відповідає комплексному спряженню. Вираз (3.3) називають *прямим перетворенням Фур'є неперервного періодичного сигналу*.

Це перетворення дає представлення сигналу $x(t)$ у вигляді коефіцієнтів його розкладу в базисі Фур'є. Набір коефіцієнтів розкладу сигналу в базисі Фур'є є проєкціями вектора, який описує сигнал, на вектори базису.

З використанням формул Ейлера можна записати ще один вираз для коефіцієнта розкладу c_n (3.3). Цей коефіцієнт є в загальному випадку комплексним числом, яке можна представити у вигляді суми дійсної та уявної частин:

$$c_n = c_{cn} + jc_{sn}.$$

Дійсна (косинусна) та уявна (синусна) частини коефіцієнту розкладу визначаються за формулами:

$$\begin{aligned} c_{cn} &= \frac{1}{T} \int_{-T/2}^{T/2} x(t) \cdot \cos(n\omega_0 t) dt, \\ c_{sn} &= \frac{1}{T} \int_{-T/2}^{T/2} x(t) \cdot \sin(n\omega_0 t) dt. \end{aligned} \quad (3.4)$$

Комплексні коефіцієнти розкладу c_n можна представити у показниковій формі:

$$c_n = |c_n| e^{j\varphi_n}, \quad (3.5)$$

де

$$\begin{aligned} |c_n| &= \sqrt{c_{cn}^2 + c_{sn}^2}, \\ \varphi_n &= \operatorname{arctg} \frac{c_{sn}}{c_{cn}}. \end{aligned} \quad (3.6)$$

З формул видно, що модуль $|c_n|$ коефіцієнта c_n є парною функцією номера n , а аргумент φ_n – непарною, оскільки функція арктангенса є непарною.

Підставивши (3.5) у (3.2), отримаємо загальний вираз розкладу сигналу у ряд Фур'є, представлений у комплексній формі:

$$x(t) = \sum_{n=-\infty}^{+\infty} |c_n| e^{j\varphi_n} e^{jn\omega_0 t} = \sum_{n=-\infty}^{+\infty} |c_n| e^{j(n\omega_0 t + \varphi_n)}. \quad (3.7)$$

Ця показникова форма ряду Фур'є для неперервної періодичної функції $x(t)$. Коефіцієнти цього ряду можна визначити за (3.3) або (3.4)-(3.6).

Тепер перейдемо до тригонометричної форми ряду Фур'є.

Зафіксуємо деякий номер $n = k$, та розглянемо пару доданків у ряді (3.7) виду $|c_n| e^{j(n\omega_0 t + \varphi_n)}$ з номерами, однаковими за модулем, але які мають протилежні знаки $n = \pm k$:

$$|c_{-k}| e^{j(-k\omega_0 t + \varphi_{-k})} + |c_k| e^{j(k\omega_0 t + \varphi_k)}.$$

Врахуємо, що фаза φ_n є непарною функцією частоти, а також застосуємо формулу Ейлера:

$$\begin{aligned}
& |c_k| \left(e^{j(-k\omega_0 t - \varphi_k)} + e^{j(k\omega_0 t + \varphi_k)} \right) = \\
& = |c_k| \left(e^{-j(k\omega_0 t + \varphi_k)} + e^{j(k\omega_0 t + \varphi_k)} \right) = \\
& = |c_k| \left(\cos(k\omega_0 t + \varphi_k) - j \sin(k\omega_0 t + \varphi_k) + \cos(k\omega_0 t + \varphi_k) + j \sin(k\omega_0 t + \varphi_k) \right) = \\
& = 2|c_k| \cos(k\omega_0 t + \varphi_k).
\end{aligned}$$

Такі розрахунки можна зробити для всіх пар доданків у (3.7), але залишиться один доданок з номером $n = 0$, для якого немає пари. Він буде мати вигляд $c_0 = |c_0| e^{j\varphi_0}$. В загальному випадку це – деяка константа. Отже, можна остаточно записати

$$x(t) = c_0 + \sum_{n=1}^{+\infty} 2|c_n| \cos(n\omega_0 t + \varphi_n). \quad (3.8)$$

Видно, що в результаті розкладу в ряд Фур'є неперервний періодичний сигнал $x(t)$ представлений сумою складових вигляду $A_n \cos(n\omega_0 t + \varphi_n)$, кожна з яких є косинусоїдальним коливанням з амплітудою A_n , частотою $n\omega_0$ та початковою фазою φ_n . Частоти елементарних коливань, з яких складено періодичну функцію $x(t)$, утворюють *гармонічну послідовність*, а саме: частоти всіх складових кратні основній частоті $\omega_0 = \frac{2\pi}{T_0}$, де T_0 – період косинусоїдального коливання найменшої (основної) частоти. Окремі складові називаються *гармоніками*. Частоти гармонік можна розрахувати за допомогою виразу:

$$\omega_n = n\omega_0 = n \frac{2\pi}{T_0}.$$

Період T_0 складової розкладу, яка має частоту ω_0 , дорівнює періоду того періодичного коливання $x(t)$, яке ми хочемо розкласти на гармоніки. Всі інші складові розкладу матимуть періоди, які менші за період T_0 в ціле число разів.

Доданок c_0 в розкладі можна розглядати як складову розкладу, частота якої нескінченно мала (дорівнює нулю). Він постійний протягом періоду T_0 , а отже описує постійну складову сигналу $x(t)$, яка не може бути представлена одним з гармонічних коливань. Величина c_0 дорівнює середньому значенню сигналу за один період (3.3).

Часто розклад сигналу в ряд Фур'є записують у такій формі:

$$\begin{aligned}
x(t) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)) = \\
&= \frac{a_0}{2} + \sum_{n=1}^{\infty} A_n \cos(n\omega_0 t + \varphi_n)
\end{aligned}$$

Видно, що для цього запису амплітуда гармоніки $A_n = 2|c_n| = \sqrt{a_n^2 + b_n^2}$, фаза $\varphi_n = \arctg \frac{b_n}{a_n}$, а відповідні коефіцієнти визначаються за формулами:

$$a_n = 2c_{cn} = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos(n\omega_0 t) dt,$$

$$b_n = 2c_{sn} = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin(n\omega_0 t) dt.$$

Отже, в результаті розкладу в ряд Фур'є ми отримали представлення періодичного сигналу у вигляді зваженої суми простих гармонічних коливань (3.8).

Сукупність значень коефіцієнтів розкладу сигналу (3.5) називають *комплексним спектром сигналу*. Структуру спектру (та властивості сигналу) повністю визначають дві характеристики спектру – амплітудна та фазова.

Амплітудна характеристика спектру сигналу – сукупність значень амплітуд гармонічних складових розкладу сигналу. Значення амплітуд дорівнюють модулю комплексного коефіцієнта розкладу.

Фазова характеристика спектру сигналу – сукупність значень початкових фаз гармонічних складових розкладу сигналу. Значення початкових фаз дорівнює аргументу комплексного коефіцієнта розкладу.

Наочне представлення про спектр кожного сигналу та його властивості можна отримати з допомогою графічного зображення амплітудного та фазового спектрів (рис. 3.2). На графіку спектру по вісі абсцис відкладається частота гармонічної складової. Оскільки частоти гармонік можуть бути лише цілими числами, то амплітудний спектр визначений не на всій числовій вісі. По вісі ординат відкладають амплітуди або початкові фази відповідних косинусоїдальних коливань.

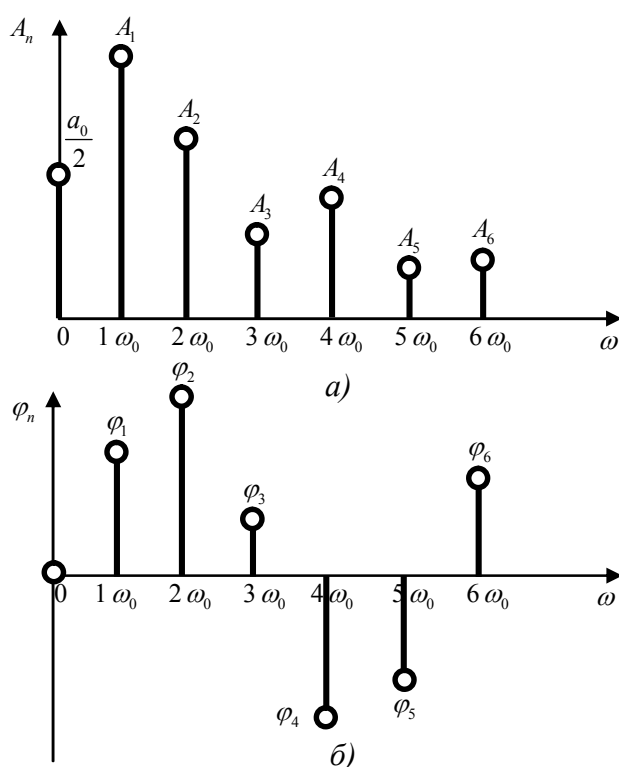


Рисунок 3.2 – а) амплітудний спектр, б) фазовий спектр

З формул та графіків видно, що *спектр неперервного періодичного нескінченного сигналу є дискретним (лінійчатим)*, тобто складається з дискретної множини точок. Кожна спектральна складова зображується своїм значенням амплітуди і початкової фази для частот, кратних основній частоті. На графіку спектрів ці числа відповідають окремим спектральним «лініям», що відповідають певним частотам.

Відстань між сусідніми спектральними лініями дорівнює *основній частоті* $\omega_0 = \frac{2\pi}{T_0}$

рад/с. Для інших значень частоти (між тими, які кратні основній) частотних складових в розкладі сигналу не існує. Якщо необхідно отримати спектральні складові на інших частотах, то єдиний шлях – змінити тривалість аналізованого сигналу (тобто період, на якому він заданий).

Треба ще раз наголосити, що в ряд Фур'є можна розкласти лише періодичні сигнали, оскільки розклад ведеться на періодичні складові: синусоїдальні та косинусоїдальні коливання кратних частот. Також сигнал, щоб його можна було представити у вигляді ряду Фур'є, має бути нескінченним. Таких сигналів не існує в природі, отже, на практиці застосувати розклад в ряд Фур'є неможливо. Але ряд Фур'є є корисною моделлю для аналізу сигналів, і в подальших розділах ми будемо робити уточнення гармонічного аналізу сигналів для представлення реальних сигналів.

Розклад сигналів на періодичні синусоїдальні компоненти і подальший аналіз результатів такого розкладу – амплітудного та фазового спектрів сигналу, займає

чільне місце у всій теорії сигналів. Майже всі системи дослідження сигналів отримують та аналізують спектральну інформацію, і розклад за Фур'є можна віднести до класичних методів теорії сигналів. Це пов'язано з тим, що багато процесів в реальному світі мають періодичний характер, і сигнали, які несуть інформацію про ці процеси, теж часто мають явні або скриті періодичні складові. Тому математичне представлення сигналу у вигляді тригонометричних функцій може набагато допомогти досліднику виявити в сигналі ці періодичності. Експериментальне здійснення розкладу сигналів на гармонічні складові з метою практичного застосування спектральних уявлень про сигнали називають *гармонічним аналізом*. Також в теорії сигналів говорять про *спектральний аналіз* сигналів – дослідження властивостей сигналів шляхом аналізу їх спектрів.

Першим, хто використав поняття спектру, був Ісаак Ньютон в 1671 році. Але ще Піфагор у VI ст. до нашої ери встановив співвідношення між періодичними коливаннями струни, які відповідають різними звукам, та довжиною цієї струни. Він вважав, що сутність звукової гармонії виражається у числах. Він розповсюдив це уявлення також і на періодичний рух небесних тіл, описавши такий рух як «музику сфер». Найбільший вклад в спектральний аналіз у сучасному його вигляді був зроблений французьким інженером Жаном Батістом Жозефом Фур'є, який у 1822 році захистив дисертацію на тему «Аналітична теорія тепла». В ній він обґрунтував, що будь-яку довільну функцію, навіть якщо вона має кінцеву кількість розривів, можна представити у вигляді нескінченної суми синусних та косинусних складових. Ця робота вважається першою роботою з гармонічного аналізу сигналів.

Але останнім часом поняття спектру часто використовують не тільки для позначення спектру за Фур'є, а і для розкладу сигналу в інших, відмінних від тригонометричного, базисах.

Також важливість і розповсюдженість гармонічного аналізу викликана тим, що гармонічні функції є власними функціями всіх лінійних систем (див. розділ 2). А отже, відомо, що кожна гармонічна складова сигналу пройде через систему, не змінивши частоту, а зміниться лише амплітуда та фаза. Тому можна досліджувати частотні характеристики систем, аналізуючи, як проходження сигналу через систему змінює його амплітудний та фазовий спектри. І навпаки, якщо ми маємо систему з відомими частотними характеристиками, то можемо передбачити, як зміняться спектральні властивості сигналу при проходженні через неї.

3.2. Розподіл енергії в спектрі неперервних періодичних сигналів

Розглянемо представлення енергетичних характеристик сигналу у спектрі. Нехай сигнал $x(t)$ описується деякою нескінченною періодичною функцією з періодом T_0 . Будемо вважати, що сигнал є напругою, яка падає на резисторі з опором 1 Ом.

Оскільки цей сигнал визначений на всій часовій вісі, то його енергія є нескінченно великою. На практиці часто необхідно знати середню потужність періодичного сигналу за період і її розподіл між окремими гармоніками сигналу. Розглянемо тригонометричний базис Фур'є $\{\phi_n\}_{n=-\infty}^{\infty} = \{e^{jn\omega_0 t}\}_{n=-\infty}^{\infty}$. Маючи формулу для середньої потужності такого сигналу за період

$$P_{\text{сеп}} = \overline{x^2(t)} = \frac{E}{t_2 - t_1} = \frac{E}{T},$$

підставимо в неї замість енергії сигналу E її вираз для сигналу, що представлений у вигляді розкладу в ортогональному тригонометричному базисі Фур'є:

$$P_{\text{сеп}} = \frac{\sum_{n=-\infty}^{+\infty} |c_n|^2 \|\phi_n\|^2}{T} = \frac{1}{T} \sum_{n=-\infty}^{+\infty} |c_n|^2 \sqrt{T}^2 = \sum_{n=-\infty}^{+\infty} |c_n|^2. \quad (3.9)$$

В разі використання тригонометричної форми ряду Фур'є можна отримати

$$P_{\text{сеп}} = \left(\frac{a_0}{2}\right)^2 + \sum_{n=1}^{\infty} A_n^2. \quad (3.10)$$

Отже, середня потужність за період дорівнює сумі квадратів коефіцієнтів розкладу сигналу в ортогональному базисі. Якщо розглядати коефіцієнти A_n як амплітуди гармонічних коливань, які складають сигнал $x(t)$, то можна сказати, що середня потужність, яка виділяється всім сигналом, дорівнює сумі потужностей, які виділяються окремо постійною складовою $\frac{a_0}{2}$ та гармоніками з амплітудами

A_1, A_2, \dots . Варто звернути увагу, у виразі для потужності (3.10) відсутня фаза гармоніки, оскільки норма базисної функції не залежить від фази комплексної експоненціальної функції. Це говорить про те, що середня потужність, яку несе сигнал за один період, не залежить від значення початкових фаз окремих гармонічних складових.

Сукупність значень квадратів коефіцієнтів розкладу сигналу в ряд Фур'є $|c_n|^2$ називається *спектром потужності* і може також бути представлена на графіку.

3.3. Аналіз неперервних неперіодичних сигналів

В попередньому підрозділі було розглянуто випадок, коли сигнал, що аналізувався, був нескінченним періодичним сигналом. Такі сигнали є корисною математичною моделлю, але достатньо рідко зустрічаються на практиці. Спектральні уявлення про сигнали можна застосувати не тільки до періодичних сигналів, але і розповсюдити

на неперіодичні неперервні сигнали, які є ближчою до реального життя моделлю сигналів. Часто ці сигнали є скінченними в часі.

Нехай маємо сигнал $x(t)$, який відмінний від нуля лише на певному проміжку $[t_1, t_2]$. Такими є всі сигнали, що вимірюються протягом скінченного інтервалу часу. Отримаємо спектральний розклад за Фур'є такого скінченного сигналу. Для цього скористуємося вже відомими формулами для розкладу періодичного сигналу в ряд Фур'є.

Виділимо довільний інтервал часу T_0 , який містить собі проміжок $[t_1, t_2]$ та представимо собі, що неперіодичний однократний сигнал $x(t)$, який заданий лише на проміжку $[t_1, t_2]$, лежить в середині періоду деякого уявного періодичного сигналу $X(t)$ з періодом T_0 . Представимо цей періодичний сигнал на графіку (рис. 3.3). Кожний період сигналу $X(t)$ містить копію сигналу $x(t)$, а на проміжку $[t_1, t_2]$ сигнали $X(t)$ та $x(t)$ співпадають.

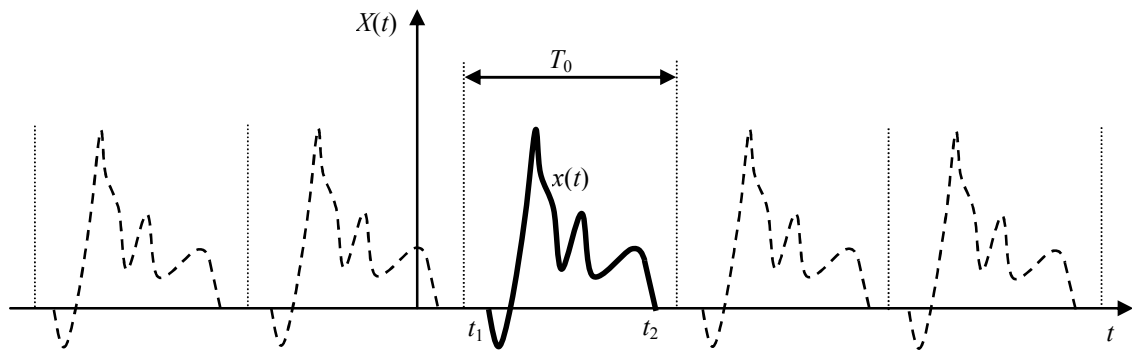


Рисунок 3.3

Оскільки отриманий сигнал $X(t)$ є періодичним з періодом T_0 нескінченним неперервним сигналом, він може бути представлений своїм рядом Фур'є:

$$X(t) = \sum_{n=-\infty}^{+\infty} c_n e^{jn\omega_0 t}, \quad (3.11)$$

де $\omega_0 = \frac{2\pi}{T_0}$ – основна частота для отриманого сигналу, а коефіцієнти розкладу визначаються за формулою:

$$c_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} X(t) e^{-jn\omega_0 t} dt. \quad (3.12)$$

Зауважимо, що в даному випадку вираз (3.11) визначає весь періодичний нескінченний сигнал $X(t)$, що простягається у часі від $-\infty$ до $+\infty$, але інтеграл береться по проміжку довжиною T_0 . Але на одному з періодів (там, де ми обрали проміжок $[t_1, t_2]$) він визначає також і наш початковий скінченний сигнал $x(t)$.

Підставимо c_n у формулу (3.11) та отримаємо вираз для нашого модельного сигналу $X(t)$ через його коефіцієнти розкладу:

$$X(t) = \sum_{n=-\infty}^{+\infty} \left(\frac{\omega_0}{2\pi} \int_{T_0} X(t) e^{-jn\omega_0 t} dt \right) e^{jn\omega_0 t}. \quad (3.13)$$

Поза проміжком T_0 ряд (3.13) визначає періодичну функцію, яка отримана повторенням $x(t)$ вліво та вправо з періодом T_0 . Треба зробити перехід від функції $X(t)$ до $x(t)$.

Ми аналізуємо функцію $x(t)$, яка є фінітною (скінченною в часі), і задана лише на інтервалі $[t_1, t_2]$. Отже, треба наблизити властивості $X(t)$ до цього сигналу $x(t)$. Для того, щоб поза відрізком $[t_1, t_2]$ функція $X(t)$ дорівнювала нулю, величина періоду T_0 повинна бути нескінченно великою. Тоді періоди функції $X(t)$ будуть відстояти один від одного на нескінченно велику відстань, і ми матимемо, по-суті, лише один нескінченно великий період уявного періодичного сигналу.

Але чим більший відрізок T_0 , який обрано в якості періоду повторення, тим меншими є коефіцієнти c_n у (3.12), оскільки множник перед інтегралом обернено пропорційний періоду. Якщо ми спрямуємо T_0 до нескінченності, то отримаємо нескінченно малі амплітуди цих гармонічних складових, сума яких складатиме вже початковий неперіодичний сигнал $x(t)$, який ми аналізували на початку. Кількість гармонічних складових ряду Фур'є буде при цьому нескінченно великою, оскільки при $T_0 \rightarrow \infty$ основна частота $\omega_0 = \frac{2\pi}{T_0} \rightarrow 0$. Тому відстань між спектральними

лініями, яка дорівнює основній частоті, стає нескінченно малою, а спектр такого сигналу стає *суцільним і неперервним*.

Отже, ми можемо зробити граничний перехід у виразі (3.13): замінити ω_0 на деяку малу частоту $d\omega$, значення частот $n\omega_0$, кратних основній частоті, – на поточну частоту ω , яка буде тепер змінюватися неперервно, а не дискретно; операцію додавання нескінченно великої кількості нескінченно малих доданків замінимо на операцію інтегрування:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} \left(\int_{t_1}^{t_2} x(t) e^{-j\omega t} dt \right) d\omega. \quad (3.14)$$

Цей вираз показує, як можна представити неперіодичний неперервний сигнал за допомогою його розкладу на гармонічні складові. Внутрішній інтеграл, який є комплексною функцією частоти:

$$F(j\omega) = \int_{t_1}^{t_2} f(t) e^{-j\omega t} dt,$$

називається *спектральною густиною* (спектральною щільністю) сигналу $x(t)$.

В загальному випадку, коли межі $[t_1, t_2]$ не визначені, спектральна густина сигналу записується для нескінченних меж

$$F(j\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt. \quad (3.15)$$

Після підстановки виразу (3.15) у (3.14) отримаємо:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\omega) e^{j\omega t} d\omega \quad (3.16)$$

Вирази (3.15) та (3.16) є відповідно *прямим та оберненим інтегральним перетворенням Фур'є для неперервних неперіодичних сигналів*.

Спектральна густина сигналу (3.15) є комплексною неперервною функцією частоти і може бути записана у показниковій формі:

$$F(j\omega) = A(\omega) - jB(\omega) = |F(j\omega)| e^{j\Phi(\omega)} = K(\omega) e^{j\Phi(\omega)}, \quad (3.17)$$

де

$$\begin{aligned} A(\omega) &= \int_{-\infty}^{+\infty} x(t) \cos(\omega t) dt, \\ B(\omega) &= \int_{-\infty}^{+\infty} x(t) \sin(\omega t) dt. \end{aligned} \quad (3.18)$$

Модуль та аргумент спектральної густини визначаються виразами

$$\begin{aligned} K(\omega) &= \sqrt{[A(\omega)]^2 + [B(\omega)]^2}, \\ \Phi(\omega) &= \arctg \frac{B(\omega)}{A(\omega)}. \end{aligned} \quad (3.19)$$

та можуть розглядатися як *амплітудно-частотна характеристика (АЧХ) спектру неперіодичного неперервного сигналу* (або *амплітудний спектр сигналу*) та *фазо-частотна характеристика (ФЧХ) спектру неперіодичного неперервного сигналу* (або *фазовий спектр сигналу*).

Варто відмітити, що між неперервним сигналом та його спектром за Фур'є існує однозначна відповідність. Значення сигналу у часі та його спектральна характеристика характеризують різні властивості одного і того самого об'єкту – сигналу, але з різних боків.

Розглянемо розмірності лівої та правої частин виразу (3.16) для визначення розмірності спектральної густини сигналу. В лівій частині сигнал має розмірність вольт (В), а в правій – добуток спектральної густини на частоту. Отже, з урахуванням (3.17) можна визначити, що модуль спектральної густини $K(\omega)$ має розмірність В/Гц або В/(рад/с). Звідси стає зрозумілим поняття спектральної «густини»: значення $K(\omega)$ показує питому амплітуду спектральних складових, що приходить на одиничний інтервал частот.

З аналізу виразів (3.19) видно, що амплітудно-частотна характеристика є парною функцією частоти, а фазо-частотна – непарною (рис. 3.4).

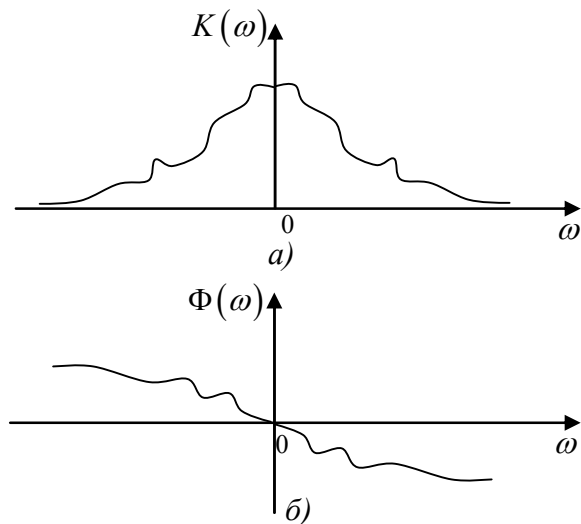


Рисунок 3.4 – а) амплітудний спектр, б) фазовий спектр

3.4. Властивості перетворення Фур'є неперервних сигналів

Розглянемо деякі властивості перетворення Фур'є неперервних сигналів. Нехай сигнал $x(t)$ має спектральну густину $F_x(j\omega)$.

1. Зсув сигналу у часі

Нехай сигнал $y(t) = x(t - t_0)$, тобто дорівнює затриманому на час t_0 сигналу $x(t)$. Спектральна густина буде записана:

$$F_y(j\omega) = \int_{-\infty}^{+\infty} x(t - t_0) e^{-j\omega t} dt = e^{-j\omega t_0} F_x(j\omega).$$

Зробимо заміну $t - t_0 = p$, тоді $t = p + t_0$, а $dt = dp$.

$$\begin{aligned} F_y(j\omega) &= \int_{-\infty}^{+\infty} x(t - t_0) e^{-j\omega t} dt = \int_{-\infty}^{+\infty} x(p) e^{-j\omega(p+t_0)} dp = \\ &= \int_{-\infty}^{+\infty} x(p) e^{-j\omega p} e^{-j\omega t_0} dp = e^{-j\omega t_0} \int_{-\infty}^{+\infty} x(p) e^{-j\omega p} dp. \end{aligned}$$

Отже, остаточно можемо записати, що

$$F_y(j\omega) = e^{-j\omega t_0} F_x(j\omega),$$

Тобто спектр зміщеного в часі сигналу дорівнює спектру початкового сигналу, який помножений на $e^{-j\omega t_0}$. Якщо представити спектральну густину $F_x(j\omega)$ у показниковій формі як $F_x(j\omega) = K_x(\omega)e^{j\Phi_x(\omega)}$, то видно, що зміщення сигналу на будь-який проміжок часу призводить до зміни лише ФЧХ сигналу:

$$F_y(j\omega) = e^{-j\omega t_0} F_x(j\omega) = e^{-j\omega t_0} K_x(\omega) e^{j\Phi_x(\omega)} = K_x(\omega) e^{j(\Phi_x(\omega) - \omega t_0)}.$$

ФЧХ зміщеного в часі сигналу набуває додаткової складової: $\Phi_y(\omega) = \Phi_x(\omega) - \omega t_0$.

Амплітудно-частотна характеристика спектру сигналу при зміщенні його в часі не змінюється, а отже, амплітудний спектр сигналу не залежить від того, де саме сигнал знаходиться на часовій вісі. Також не змінюється характер розподілу енергії гармонічних складових по частоті.

2. Множення на комплексний експоненціальний сигнал

Знайдемо спектр сигналу $y(t) = x(t) \cdot e^{j\omega_0 t}$:

$$F_y(j\omega) = \int_{-\infty}^{+\infty} x(t) e^{j\omega_0 t} e^{-j\omega t} dt = \int_{-\infty}^{+\infty} x(t) e^{-j(\omega - \omega_0)t} dt.$$

Отриманий вираз є прямим перетворенням Фур'є для частот, які зміщені по частотній вісі праворуч на ω_0 :

$$F_y(j\omega) = F_x(j(\omega - \omega_0)),$$

Отже, множення сигналу на комплексний експоненціальний сигнал частоти ω_0 в часовій області призводить до того, що його спектр зміщується по вісі частот на ω_0 .

3. Множення на косинусоїдальний сигнал

Нехай $y(t) = x(t) \cdot \cos(\omega_0 t)$, тоді

$$\begin{aligned} F_y(j\omega) &= \int_{-\infty}^{+\infty} x(t) \cos(\omega_0 t) e^{-j\omega t} dt = \int_{-\infty}^{+\infty} x(t) \frac{1}{2} (e^{j\omega_0 t} + e^{-j\omega_0 t}) e^{-j\omega t} dt = \\ &= \frac{1}{2} \left(\int_{-\infty}^{+\infty} x(t) e^{j\omega_0 t} e^{-j\omega t} dt + \int_{-\infty}^{+\infty} x(t) e^{-j\omega_0 t} e^{-j\omega t} dt \right) = \\ &= \frac{1}{2} F_x(j(\omega - \omega_0)) + \frac{1}{2} F_x(j(\omega + \omega_0)). \end{aligned}$$

Отже, при множенні сигналу на косинусоїдальний сигнал частоти ω_0 , спектр отриманого сигналу буде складатися із двох частин-копій спектру початкового сигналу. Перша частина буде зміщена на величину ω_0 ліворуч відносно початкового спектру, а друга – на ω_0 праворуч від початкового спектру. По амплітуді всі

спектральні складові будуть вдвічі меншими від їх значень у спектрі початкового сигналу.

4. Відзеркалення сигналу в часовій області

Нехай маємо сигнал, який є затриманим в часі і розгорнутим по часовій вісі відносно сигналу $x(t)$: $y(t) = x(t_0 - t)$.

$$F_y(j\omega) = \int_{-\infty}^{+\infty} x(t_0 - t) e^{-j\omega t} dt.$$

Зробимо заміну змінних: $t_0 - t = p$, тоді $t = t_0 - p$, а $dt = -dp$. Також зміняться межі інтегрування. Повернемо їх до початкового стану, прибравши знак «мінус» перед інтегралом. При цьому

$$\begin{aligned} F_y(j\omega) &= - \int_{+\infty}^{-\infty} x(p) e^{-j\omega(t_0 - p)} dp = e^{-j\omega t_0} \int_{-\infty}^{+\infty} x(p) e^{j\omega p} dp = \\ &= e^{-j\omega t_0} F_x^*(j\omega). \end{aligned}$$

Отже, при оберненні сигналу в часі та його затримці спектр множиться на функцію $e^{-j\omega t_0}$, та стає комплексно-спряженим відносно спектру початкового сигналу.

5. Розтяг або стиснення сигналу

Нехай початковий сигнал $x(t)$ піддали розтягу чи стисненню з коефіцієнтом a : $y(t) = x(at)$. Якщо величина $a > 1$, то $y(t)$ буде стиснутим в часі відносно $x(t)$, а якщо $a < 1$ – розтягнутим по часовій вісі. Спектральна густина стиснутого або розтягнутого сигналу дорівнює:

$$F_y(j\omega) = \int_{-\infty}^{+\infty} x(at) e^{-j\omega t} dt = \frac{1}{a} F_x\left(\frac{\omega}{a}\right).$$

Отже, при стисненні сигналу у часі, в стільки ж разів розтягується його спектр по вісі частот з одночасним зменшенням амплітуди спектральної функції. І навпаки, якщо сигнал у часі розтягується, то його спектр стає більш вузьким та високоамплітудним.

6. Спектр суми сигналів

Нехай сигнал $x(t)$ має спектральну густину $F_x(j\omega)$, а сигнал $y(t)$ – спектральну густину $F_y(j\omega)$. Оскільки перетворення Фур'є (3.15) є лінійним перетворенням, то при додаванні сигналів додаються і їх спектральні густини: сигналу $ax(t) + by(t)$ буде відповідати спектральна густина $aF_x(j\omega) + bF_y(j\omega)$.

7. Спектр добутку двох сигналів

Нехай сигнал $y(t) = x(t)z(t)$. Запишемо вираз для спектру такого добутку, причому один з сигналів-множників представимо через його обернене перетворення Фур'є:

$$\begin{aligned}
 F_y(j\omega) &= \int_{-\infty}^{+\infty} x(t)z(t)e^{-j\omega t} dt = \\
 &= \int_{-\infty}^{+\infty} x(t) \underbrace{\frac{1}{2\pi} \int_{-\infty}^{+\infty} F_z(j\nu) e^{j\nu t} d\nu}_{z(t)} e^{-j\omega t} dt = \\
 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_z(j\omega) \underbrace{\int_{-\infty}^{+\infty} x(t) e^{-j(\omega-\nu)t} dt}_{F_x(j(\omega-\nu))} d\nu = \int_{-\infty}^{+\infty} F_z(j\omega) F_x(j(\omega-\nu)) d\nu.
 \end{aligned}$$

Отже, спектр добутку двох сигналів буде дорівнювати згортці спектрів двох сигналів, помноженій на коефіцієнт $\frac{1}{2\pi}$:

$$F_y(j\omega) = \frac{1}{2\pi} F_x(j\omega) * F_z(j\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_z(j\omega) F_x(j(\omega-\nu)) d\nu.$$

8. Спектр згортки двох сигналів

Нехай сигнал $y(t) = x(t) * z(t) = \int_{-\infty}^{+\infty} x(\tau)z(t-\tau) d\tau$. Знайдемо його спектр:

$$\begin{aligned}
 F_y(j\omega) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x(\tau)z(t-\tau)e^{-j\omega t} dt d\tau = \\
 &= \int_{-\infty}^{+\infty} x(\tau) \left[\int_{-\infty}^{+\infty} z(t-\tau)e^{-j\omega t} dt \right] d\tau = \\
 &= \int_{-\infty}^{+\infty} x(\tau) \left[e^{j\omega\tau} F_z(j\omega) \right] d\tau = F_z(j\omega) \int_{-\infty}^{+\infty} x(\tau) e^{j\omega\tau} d\tau = \\
 &= F_x(j\omega) F_z(j\omega).
 \end{aligned}$$

Отже, спектр згортки сигналів буде дорівнювати добутку спектрів:

$$F_y(j\omega) = F_x(j\omega) F_z(j\omega).$$

З цього випливає, що результат згортки двох сигналів в часі можна визначити в спектральній області через обернене перетворення Фур'є:

$$s(t) = \int_{-\infty}^{+\infty} x(\tau)y(t-\tau)d\tau = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_x(j\omega)F_y(j\omega)e^{j\omega t}d\omega.$$

Останній вираз широко використовується в теорії лінійних систем при аналізі проходження сигналів через лінійні кола. В цьому випадку для розрахунку спектру вихідного сигналу достатньо розрахувати добуток спектру вхідного сигналу на комплексну частотну характеристику.

9. Спектр похідної від сигналу

Нехай $y(t) = \frac{dx(t)}{dt}$, тоді

$$F_y(j\omega) = \int_{-\infty}^{+\infty} \frac{dx(t)}{dt} e^{-j\omega t} dt.$$

Скористаємось визначенням похідної $\frac{dx(t)}{dt} = \lim_{\tau \rightarrow 0} \frac{x(t) - x(t-\tau)}{\tau}$ та підставимо його під знак інтегралу:

$$\begin{aligned} F_y(j\omega) &= \lim_{\tau \rightarrow 0} \int_{-\infty}^{+\infty} \frac{x(t) - x(t-\tau)}{\tau} e^{-j\omega t} dt = \\ &= \lim_{\tau \rightarrow 0} \frac{\int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt - \int_{-\infty}^{+\infty} x(t-\tau) e^{-j\omega t} dt}{\tau} = \\ &= \lim_{\tau \rightarrow 0} \frac{F_x(j\omega) - e^{-j\omega\tau} F_x(j\omega)}{\tau} = F_x(j\omega) \lim_{\tau \rightarrow 0} \frac{1 - e^{-j\omega\tau}}{\tau}. \end{aligned}$$

Розкладемо $e^{-j\omega\tau}$ в ряд при умові, що значення τ прямує до нуля, і обмежимося лише першими двома членами цього ряду:

$$e^{-j\omega\tau} = 1 - j\omega\tau + \frac{(\omega\tau)^2}{2} - \dots = 1 - j\omega\tau.$$

Остаточно при граничному переході отримаємо:

$$F_y(j\omega) = j\omega F_x(j\omega).$$

Отже, диференціювання в часовій області еквівалентне множенню на $j\omega$ в частотній області.

10. Спектр інтеграла від сигналу

Якщо сигнал $y(t)$ є похідною від $x(t)$, то $x(t)$ – є інтегралом від $y(t)$. Використовуючи попередню властивість, можна записати, що спектр інтегралу від сигналу $z(t) = \int x(t)$ буде дорівнювати:

$$F_z(j\omega) = \frac{F_x(j\omega)}{j\omega}.$$

3.5. Розподіл енергії в спектрі неперіодичного сигналу

Розглянемо випадок, коли треба отримати спектр добутку сигналу самого на себе, тобто коли $s(t) = x(t)x(t) = x^2(t)$. Для цього підставимо в **Ошибка! Источник ссылки не найден.** значення $\Omega = 0$ та розрахуємо спектральну густину:

$$\begin{aligned} F_s(j0) &= \int_{-\infty}^{+\infty} x^2(t) e^{-j0t} dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_x(j\omega) F_x(-j\omega) d\omega = \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_x(j\omega) F_x^*(j\omega) d\omega, \end{aligned} \quad (3.20)$$

де $F_x^*(j\omega)$ – функція, комплексно спряжена для функції $F_x(j\omega)$.

Якщо сигнал $x(t)$ описує електричну величину (напругу або струм), то інтеграл

$$\int_{-\infty}^{+\infty} x^2(t) dt$$

представляє собою повну енергію сигналу $x(t)$. Добуток спектральних густин $F_x(j\omega) F_x^*(j\omega)$ дорівнює $|F_x(j\omega)|^2$, тобто квадрату модуля спектральної густини. Отже, повна енергія сигналу дорівнює:

$$E = \int_{-\infty}^{+\infty} x^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |F_x(j\omega)|^2 d\omega. \quad (3.21)$$

Цей вираз є рівністю Парсеваля для випадку неперервних неперіодичних сигналів та показує зв'язок між енергетичними характеристиками та модулем спектральної густини сигналу. Видно, що повна енергія неперервного неперіодичного сигналу при його спектральному розкладі перерозподіляється між енергією всіх гармонічних складових.

Якщо розглянути розмірність величин у виразі (3.21), то можна сказати, що величина квадрату модуля спектральної густини (амплітудно-частотної характеристики спектру) має сенс спектральної густини енергії сигналу, і показує, яка енергія приходить на одиницю смуги частот в спектрі сигналу. Одиниці вимірювання спектральної густини потужності – В²/Гц.

3.6. Аналіз дискретних нескінченних сигналів

До цього часу ми розглядали задачу спектрального аналізу за Фур'є лише для неперервних сигналів. Але для практичного застосування більш важливим є можливість розрахунку спектральних характеристик для дискретизованих сигналів, оскільки переважна більшість систем аналізу сигналів працює з дискретними сигналами.

Для отримання виразів амплітудно-частотної та фазо-частотної характеристики спектрів дискретних сигналів, перейдемо від спектральних характеристик неперервного сигналу до дискретного. По-перше, згадаємо представлення дискретного сигналу з допомогою послідовності зміщених одиничних імпульсів. Нехай маємо неперервний сигнал $x(t)$ нескінченної тривалості, який дискретизуємо з періодом T_s . Використаємо функцію Кронекера

$$\delta(t) = \begin{cases} 1, & t = 0, \\ 0, & t \neq 0 \end{cases}$$

і запишемо вираз для результуючого дискретного сигналу $x[n]$:

$$x[n] = \sum_{n=-\infty}^{+\infty} x(t) \delta(t - nT_s). \quad (3.22)$$

Можна також записати еквівалентні вирази:

$$x[n] = \sum_{n=-\infty}^{+\infty} x(nT_s) \delta(t - nT_s) = x(t) \sum_{n=-\infty}^{+\infty} \delta(t - nT_s).$$

Права частина виразу (3.22) може розглядатися як неперервний сигнал, оскільки представляє собою набір функцій Кронекера, кожна з яких розміщена у момент nT_s , в який береться відлік неперервного сигналу, і має величину, що рівна значенню сигналу $x(t)$ в цей момент. Будемо користуватися формулою для інтегрального перетворення Фур'є (3.15), а не для ряду Фур'є, оскільки неперервний сигнал $x(t)$, який піддається дискретизації, в загальному випадку може бути неперіодичним. Підставимо у (3.15) праву частину виразу (3.22). Можна записати такі перетворення:

$$\begin{aligned}
F(j\omega) &= \int_{-\infty}^{+\infty} x[n] e^{-j\omega t} dt = \\
&= \int_{-\infty}^{+\infty} \left(\sum_{n=-\infty}^{+\infty} x(t) \delta(t - nT_s) \right) e^{-j\omega t} dt = \\
&= \sum_{n=-\infty}^{+\infty} \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} \delta(t - nT_s) dt = \\
&= \sum_{n=-\infty}^{+\infty} x(nT_s) e^{-j\omega nT_s} = \\
&= \sum_{n=-\infty}^{+\infty} x[n] e^{-j\omega nT_s}.
\end{aligned} \tag{3.23}$$

Отже, отримано вираз, який дозволяє розрахувати *спектральну густину дискретного нескінченного сигналу* $x[n]$:

$$F(j\omega) = \sum_{n=-\infty}^{+\infty} x[n] e^{-j\omega nT_s}. \tag{3.24}$$

Як видно, спектральна густина дискретного нескінченного сигналу є неперервною комплексною функцією частоти. Також можна помітити, що спектр дискретного сигналу є *періодичною* неперервною функцією:

$$e^{-j\omega nT_s} = e^{-j(\omega \pm k\omega_s)nT_s} = e^{-j\left(\omega \pm k\frac{2\pi}{T_s}\right)nT_s} = e^{-j\omega nT_s} e^{\mp j2\pi kn} = e^{-j\omega nT_s},$$

причому період повторення цієї функції в частотній області дорівнює частоті дискретизації ω_s . Можна вважати, що у виразі (3.24) множник $x[n]$ перед комплексною експонентою є «коефіцієнтом» в ряді Фур'є. В цьому випадку ми користуємось дуальністю понять часу і частоти: якщо у виразах для прямого та оберненого перетворення замінити час на частоту, то вирази залишаться правильними. Отже, для отримання оберненого перетворення Фур'є (виразу для дискретного сигналу через його спектр), можна скористатися виразами рядом Фур'є (3.3), замінивши час на частоту:

$$x[n] = \frac{1}{\omega_s} \int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} F(j\omega) e^{j\omega nT_s} d\omega. \tag{3.25}$$

Вирази (3.24) та (3.25) називаються відповідно *прямим та оберненим неперервним перетворенням Фур'є дискретного нескінченного сигналу*, а величина $F(j\omega)$ – Фур'є-образом послідовності $x[n]$ (комплексним спектром нескінченного дискретного сигналу). Комплексний спектр можна записати в алгебраїчній та показниковій формах:

$$\begin{aligned}
 F(j\omega) &= \operatorname{Re}(F(j\omega)) + j \operatorname{Im}(F(j\omega)), \\
 F(j\omega) &= |F(j\omega)| e^{j \arg(F(j\omega))} = K(\omega) e^{j\Phi(\omega)}.
 \end{aligned}
 \tag{3.26}$$

Тут величини $K(\omega)$ та $\Phi(\omega)$ – відповідно амплітудний та фазовий спектр сигналу (їх іноді називають амплітудно-частотною та фазо-частотною характеристиками спектру сигналу).

Іноді для спрощення запису приймають, що крок дискретизації дорівнює одиниці, $T_s = 1$ с, і частота дискретизації дорівнює 1 Гц або 2π рад/с. Це так зване «нормування частоти». В цьому випадку вирази для прямого і оберненого перетворень Фур'є матимуть вигляд:

$$\begin{aligned}
 F(j\omega) &= \sum_{n=-\infty}^{+\infty} x[n] e^{-j\omega n}, \\
 x[n] &= \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\omega) e^{j\omega n} d\omega.
 \end{aligned}$$

3.7. Властивості перетворення Фур'є дискретних нескінченних сигналів

Існує кілька теорем, які встановлюють співвідношення між операціями над дискретними сигналами та їх спектрами. Введемо скорочені позначення: $F(j\omega) = \mathbb{F}\{x[n]\}$, $x[n] = \mathbb{F}^{-1}\{F(j\omega)\}$, $x[n] \xleftrightarrow{\mathbb{F}} F(j\omega)$.

Лінійність перетворення Фур'є

Якщо $x_1[n] \xleftrightarrow{\mathbb{F}} F_1(j\omega)$, а $x_2[n] \xleftrightarrow{\mathbb{F}} F_2(j\omega)$, то безпосередньо з означення перетворення Фур'є та властивості лінійності інтегралу випливає, що

$$ax_1[n] + bx_2[n] \xleftrightarrow{\mathbb{F}} aF_1(j\omega) + bF_2(j\omega).$$

Спектр дійсного сигналу

Для дійсного сигналу модуль його спектру є парною функцією частоти, а аргумент – непарною.

$$\begin{aligned}
 |F(j\omega)| &= |F(-j\omega)|, \\
 \arg(F(j\omega)) &= -\arg(F(-j\omega)).
 \end{aligned}$$

Часовий зсув

Якщо $x_1[n] \xleftrightarrow{\mathbb{F}} F_1(j\omega)$, то при затримці сигналу у часі на величину n_d , то його перетворення Фур'є змінюється за таким правилом:

$$\begin{aligned}\mathbb{F}\{x[n - n_d]\} &= \sum_{n=-\infty}^{+\infty} x[n - n_d] e^{-j\omega n T_s} = \sum_{k=-\infty}^{+\infty} x[k] e^{-j\omega(n - n_d + n_d) T_s} = \\ &= \sum_{n=-\infty}^{+\infty} x[n - n_d] e^{-j\omega(n - n_d) T_s} e^{-j\omega n_d T_s} = e^{-j\omega n_d T_s} \sum_{k=-\infty}^{+\infty} x[k] e^{-j\omega k T_s} = e^{-j\omega n_d T_s} F(j\omega).\end{aligned}$$

Видно, що зсув у часі функції $x[n]$ на величину $\pm n_d$ призводить до зміни фазової характеристики на $\pm \omega n_d T_s$. Амплітудна характеристика та спектральна щільність потужності від положення сигналу на часовій вісі не залежить.

Частотний зсув

Аналогічно у частотній області, якщо ми помножимо сигнал на синусоїдальний сигнал частоти ω_0 , $e^{j\omega_0 n T_s}$, то отримаємо:

$$\begin{aligned}\mathbb{F}\{e^{j\omega_0 n T_s} x[n]\} &= \sum_{n=-\infty}^{+\infty} e^{j\omega_0 n T_s} x[n] e^{-j\omega n T_s} = \\ &= \sum_{n=-\infty}^{+\infty} x[n] e^{-j(\omega - \omega_0) n T_s} = F(j(\omega - \omega_0)).\end{aligned}$$

Отже, при множенні сигналу у часовій області на коливання певної частоти, спектр сигналу зсувається на цю частоту. Ця властивість буде докладніше розглянута при обговоренні модуляції сигналів.

Диференціювання спектру

Якщо $x[n] \xleftrightarrow{\mathbb{F}} F(j\omega)$, то в разі, якщо ми візьмемо похідну від спектру, отримаємо:

$$j \frac{dF(j\omega)}{d\omega} \xleftrightarrow{\mathbb{F}} nx[n].$$

Диференціювання сигналу

Відповідає диференціюванню всіх його гармонічних складових. Похідна від $e^{j\omega t}$ дорівнює $j\omega$, а отже можна записати:

$$\frac{dx(t)}{dt} \leftrightarrow j\omega F(j\omega).$$

Рівність Парсеваля

$$E = \sum_{n=-\infty}^{+\infty} |x[n]|^2 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} |F(j\omega)|^2 d\omega.$$

Функція $|F(j\omega)|^2$ є спектральною густиною потужності дискретного сигналу та показує, як енергія сигналу розподілена у частотній області. Ця рівність встановлює зв'язок між енергією дискретного сигналу, яка обчислена в часовій та частотній областях.

Кругова (періодична) згортка в часовій області

Нехай $x[n] \xleftrightarrow{\mathbb{F}} F(j\omega)$ та $h[n] \xleftrightarrow{\mathbb{F}} H(j\omega)$. Тоді в разі, якщо

$$y[n] = \sum_{k=-\infty}^{+\infty} x[k]h[n-k] = x[n] * h[n],$$

то

$$Y(\omega) = F(\omega)H(\omega),$$

тобто згортка двох послідовностей у часовій області відповідає добутку Фурьє-образів цих послідовностей.

Множення сигналів в часовій області

Якщо $x[n] \xleftrightarrow{\mathbb{F}} F(j\omega)$ та $g[n] \xleftrightarrow{\mathbb{F}} G(\omega)$, а

$$y[n] = f[n] \cdot g[n]$$

то

$$Y(\omega) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} F(\theta)G(\omega - \theta)d\theta.$$

Отже, множення сигналів у часовій області відповідає згортці їх спектрів у частотній області. Права частина рівняння називається періодичною згорткою, тобто згорткою двох періодичних функцій, проміжок інтегрування яких співпадає з періодом підінтегральних функцій.

Масштабування сигналу

Якщо $f[n] \xleftrightarrow{\mathbb{F}} F(j\omega)$, то

$$x[an] \xleftrightarrow{\mathbb{F}} \frac{1}{|a|} F\left(j\frac{\omega}{a}\right).$$

При стисненні сигналу в a разів на часовій вісі в стільки ж разів розширюється його спектр на вісі частот, а модуль спектральної щільності при цьому зменшується в a разів.

3.8. Зв'язок між спектрами неперервного та дискретного сигналів

Важливим є питання, а що відбувається зі спектром неперервного сигналу при його дискретизації. Ці знання допоможуть нам провести процедуру аналого-цифрового перетворення так, щоб не втратити істотні властивості сигналу при перетворенні його з аналогового на цифровий. Знання ефектів та властивостей частотного представлення дискретизації дасть нам можливість досліджувати спектри сигналів реальних об'єктів на основі аналізу оцифрованих сигналів.

Нехай маємо нескінченний аналоговий сигнал $x_c(t)$ та сигнал $x[n] = x_c(nT_s)$, який отримано при його дискретизації з частотою $\omega_s = \frac{2\pi}{T_s}$. Знайдемо зв'язок між їх спектрами.

Для неперервного сигналу пряме та обернене інтегральне перетворення Фур'є можна записати:

$$F_c(\omega) = \int_{-\infty}^{+\infty} x_c(t) e^{-j\omega t} dt,$$

$$x_c(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_c(\omega) e^{j\omega t} d\omega.$$

Замінімо для сигналу $x_c(t)$ неперервний час дискретним $t \rightarrow nT_s$ та запишемо:

$$x_c(nT_s) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_c(\omega) e^{j\omega nT_s} d\omega.$$

Згадуючи властивості невизначеного інтеграла, обчислення інтегралу на інтервалі $(-\infty, +\infty)$ замінимо обчисленням нескінченної суми інтегралів на рівних інтервалах

шириною $\Delta\omega = \omega_s = \frac{2\pi}{T_s}$:

$$x_c(nT_s) = \sum_{m=-\infty}^{+\infty} \frac{1}{2\pi} \int_{m\omega_s - \frac{\omega_s}{2}}^{m\omega_s + \frac{\omega_s}{2}} F_c(\omega) e^{j\omega nT_s} d\omega,$$

де m – номер інтервалу.

Інтеграл зі змінними межами від незалежного від m підінтегрального виразу можна замінити на інтеграл з фіксованими межами, але від залежного від m підінтегрального виразу $F_c(\omega + m\omega_s)$. Цей вираз є спектром аналогового сигналу, що зміщений по вісі частот на крок $m\omega_s$, де $m = 0, \pm 1, \pm 2, \dots$. Якщо у всьому виразі замінити ω на $\omega + m\omega_s$, рівність буде мати вигляд:

$$x_c(nT_s) = \frac{1}{2\pi} \sum_{m=-\infty}^{+\infty} \int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} F_c(\omega + m\omega_s) e^{j(\omega + m\omega_s)nT_s} d(\omega + m\omega_s)$$

Можна відмітити, що:

$e^{j(\omega + m\omega_s)nT_s} = e^{j\omega nT_s}$, оскільки функція періодична;

$d(\omega + m\omega_s) = d\omega$ внаслідок властивості диференціалу.

Отже:

$$x_c(nT_s) = \frac{1}{2\pi} \sum_{m=-\infty}^{+\infty} \int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} F_c(\omega + m\omega_s) e^{j\omega nT_s} d\omega$$

Змінимо порядок інтегрування та додавання, оскільки вони проводяться по різним незалежним змінним, і остаточно отримаємо:

$$x(nT_s) = x[n] = \frac{1}{2\pi} \int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} \underbrace{\sum_{m=-\infty}^{+\infty} F_c(\omega + m\omega_s)}_{F_d(j\omega)} e^{j\omega nT_s} d\omega. \quad (3.27)$$

Порівняємо отриманий вираз з виразом для оберненого перетворення Фур'є нескінченного дискретного сигналу (3.25):

$$x[n] = \frac{1}{\omega_s} \int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} F(j\omega) e^{j\omega nT_s} d\omega = \frac{T_s}{2\pi} \int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} F(j\omega) e^{j\omega nT_s} d\omega.$$

Ліві частини виразів однакові, отже праві також мають бути однаковими. Вони відрізняються наявністю множника T_s та підінтегральними виразами. Можна сказати, що підінтегральний вираз (сума) в (3.27) є спектром за Фур'є нескінченного дискретного сигналу. Отже, можемо визначити зв'язок між спектрами неперервного $F_c(j\omega)$ та дискретного $F_d(j\omega)$ сигналів:

$$F_d(\omega) = \frac{1}{T_s} \sum_{m=-\infty}^{+\infty} F_c(\omega + m\omega_s). \quad (3.28)$$

Тобто, спектр дискретного нескінченного сигналу рівний (з точністю до множника $\frac{1}{T_s}$) сумі нескінченної кількості спектрів неперервного сигналу, які зміщені по вісі

частот на частоти, кратні частоті дискретизації: $m\omega_s = m\frac{2\pi}{T_s}$, де $\omega = 0, \pm 1, \pm 2, \dots$

Отже, спектр дискретного сигналу є нескінченна сума копій спектрів неперервного сигналу, зміщених один відносно одного на частоту дискретизації. Ми показали, що спектр Фур'є дискретного сигналу $x[n]$ складається з копій Фур'є-спектру сигналу

$x(t)$, які періодично повторюються. Період повторення екземплярів спектру дорівнює частоті дискретизації.

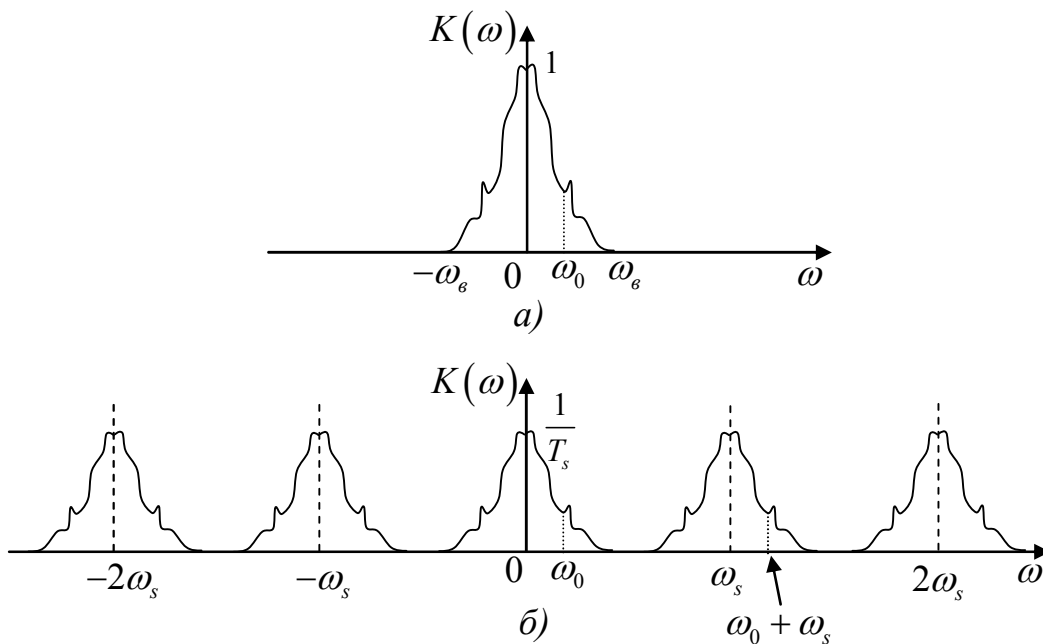


Рисунок 3.5 – а) амплітудний спектр початкового неперервного сигналу, б) амплітудний спектр того самого сигналу після дискретизації з частотою ω_s

Нехай маємо неперервний (в загальному випадку – неперіодичний) сигнал з обмеженим спектром, в якому спектральна складова з ненульовою амплітудою має максимальну частоту ω_g (рис. 3.5, а). Поза проміжком $[-\omega_g, \omega_g]$ спектральних складових немає, спектр там дорівнює нулю.

При дискретизації цього сигналу з частотою дискретизації ω_s матимемо в спектральній області періодичне продовження копій цього спектру. Спектр дискретизованого сигналу буде мати вигляд, представлений на рисунку 3.5, б: він залишається неперервною функцією частоти (оскільки функція (3.24) та (3.28) – неперервні функції), але буде тепер простягатися від $-\infty$ до $+\infty$. Головна особливість – періодизація спектру внаслідок дискретизації. Спектр буде періодичною функцією частоти. Величини складових отриманого періодичного спектру будуть в T_s разів менше, ніж в спектрі початкового неперервного сигналу.

Періодичність спектру дискретизованого сигналу – одне з головних явищ, які відбуваються внаслідок аналогово-цифрового перетворення. Якщо обрати частоти в різних періодах спектру, які будуть відрізнятися на частоту дискретизації $\omega_k = k\omega_s \pm \omega_0$, $k = 0, 1, 2, \dots$, то значення спектральних складових на цих частотах будуть однаковими.

Розглянемо простий випадок, коли ми дискретизуємо неперервний нескінченний синусоїдальний сигнал $x(t) = \sin(\omega_0 t)$. Для загальності використаємо інтегральне перетворення Фур'є, і в цьому випадку отримаємо амплітудний спектр, зображений

на рисунку 3.6, а. Він складається з двох спектральних ліній (оскільки сигнал містить лише єдину синусоїдальну складову) для частот $-\omega_0$ та ω_0 (оскільки внаслідок властивостей інтегрального перетворення Фур'є амплітудний спектр є парною функцією частоти).

Відповідно до властивості періодичності (3.28), спектр синусоїдального сигналу після дискретизацію буде мати вигляд як на рис. 3.6, б, а саме, складатися із копій спектру неперервної нескінченної синусоїди з рисунку 3.6, а.

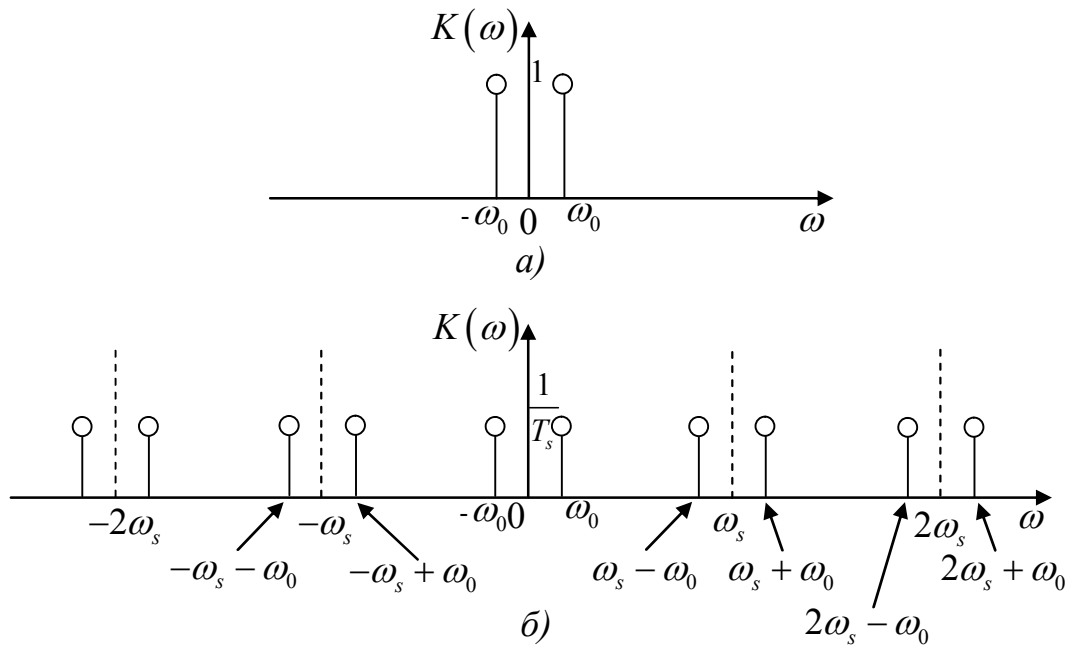


Рисунок 3.6 – а) амплітудний спектр синусоїдального неперервного сигналу частоти ω_0 , б) амплітудний спектр синусоїдального сигналу частотою ω_0 після дискретизації з частотою ω_s

Можна побачити, що спектр після дискретизації складається зі багатьох спектральних ліній, кожен з яких можна інтерпретувати як синусоїдальну складову. Отже, якби ми мали в часовій області синусоїди, частоти яких дорівнюють $k\omega_s \pm \omega_0$, $k=0,1,2,\dots$ то вони мали б такий самий спектр, як і синусоїда $x(t) = \sin(\omega_0 t)$. Дивлячись на спектр на рис. 3.6, б, не можна сказати, якою саме була частота синусоїди в часовій області. Якщо розглянути це явище в часовій області, то ця невизначеність – аналог того факту, що після дискретизації сигналу отримаємо лише набір чисел. І знаючи тільки ці числа, неможливо сказати, яка була частота дискретизації. А отже, невідомо, яка була частота синусоїди. Іншими словами, будь-яка з копій спектру на рис. 3.5, б або 3.6, б, може бути спектром реального неперервного сигналу.

Якщо повернутися до прикладу синусоїди, то розглянемо простий синусоїдальний сигнал в часі, який дискретизований з частотою ω_s і додамо до частоти цілу кількість частот дискретизації:

$$\begin{aligned}
 x[n] &= \sin(\omega_0 n T_s) = \\
 &= \sin\left(\underbrace{(\omega_0 + k\omega_s) n T_s}_{\text{частота}}\right) = \sin\left(\left(\omega_0 + k \frac{2\pi}{T_s}\right) n T_s\right) = \\
 &= \sin\left(\omega_0 n T_s + 2\pi \underbrace{kn}_m\right) = \sin(\omega_0 n T_s).
 \end{aligned}$$

З цих математичних перетворень також видно, що при дискретизації сигналу з частотою ω_s , ми не можемо розрізнити дискретизовані значення синусоїди частотою ω_0 та синусоїди частотою $\omega_0 + k\omega_s$, бо ці відліки результуючого дискретного сигналу будуть однаковими. Отримана послідовність представляє собою нескінченну кількість різних синусоїд.

Наприклад, на рисунку 3.7 наведено ділянку синусоїди частоти 1.25 Гц (штрихова лінія) та 6 Гц (суцільна лінія). Видно, що існують деякі моменти часу, в які ці синусоїди мають однакові значення. На рисунку вони позначені кружечками. Якщо частота дискретизації буд обрана такою, що відліки будуть виміряні в саме ці моменти часу, то незалежно від того, якої частоти була неперервна синусоїда, дискретний сигнал буде один и той самий.

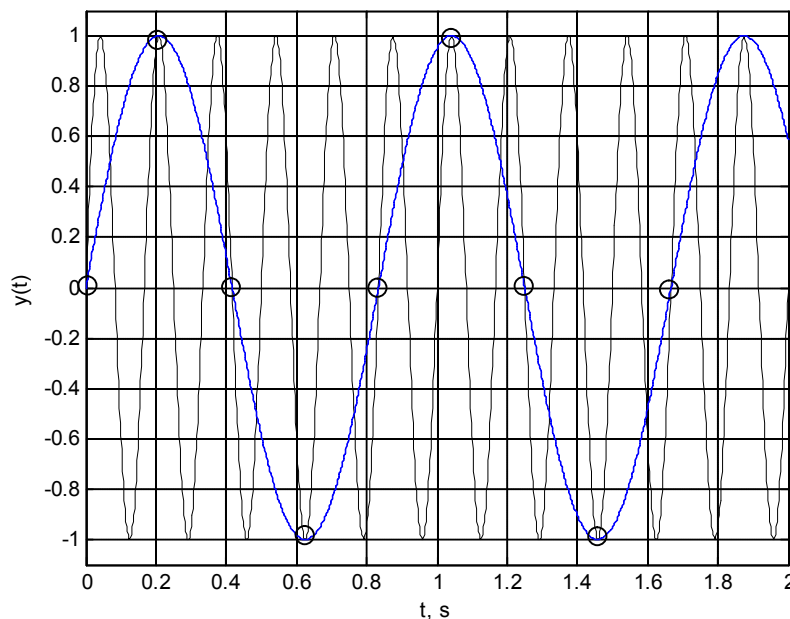


Рисунок 3.7 – Синусоїдальні сигнали різних частот, які можуть мати однакові відліки (відмічені кружечками) після дискретизації

3.9. Перетворення Фур'є фінітних дискретних сигналів

Як відомо, всі реальні цифрові сигнали, які приходить обробляти в реальному житті, не тільки дискретні, але й фінітні. Навіть якщо обробка потоку цифрових відліків сигналу, які безперервно поступають на вхід дискретної системи обробки, проводиться у реальному часі, завжди використовують буферизацію, при якій поки іде обробка одного блоку даних, в пам'яті системи накопичуються подальші відліки сигналу, які приходять в цей час на вхід. Оскільки спектральний аналіз за Фур'є є одним з основних методів дослідження даних, і всі застосування теорії сигналів працюють саме з дискретними скінченними сигналами, треба мати алгоритми спектрального аналізу для них.

Нехай після аналого-цифрового перетворення маємо N відліків сигналу $x(t)$. Математичною моделлю таких сигналів будуть вектори в просторі C^N , тобто послідовності N впорядкованих комплексних (в загальному випадку) чисел. Будемо нумерувати ці N чисел індексами $n = \{0, 1, 2, \dots, N-1\}$. Позначимо послідовність x (яка моделює наш сигнал) як функцію, яка визначена на скінченній множині точок $Z_N = \{0, 1, 2, \dots, N-1\}$:

$$x[n] = [x[0], x[1], x[2], \dots, x[N-1]].$$

Якщо ми задамо звичайним чином покомпонентне додавання та множення на скаляр, то C^N буде N -вимірним векторним гільбертовим простором. Для того, щоб отримати спектр сигналу з цього простору, треба задати в ньому аналог базису з гармонічних синусоїдальних коливань.

Задамо в цьому просторі набір дискретних функцій $F_m[n] \in C^N$:

$$F_m[n] = e^{2\pi j \frac{m}{N} n}, \quad (3.29)$$

де $n = 0, 1, 2, \dots, N-1$ – номер відліку функції;

$m = 0, 1, 2, \dots, N-1$ – порядковий номер функції.

Це набір дискретних комплексних тригонометричних функцій. Можна показати, що такий набір є базисом простору C^N . Функції $\{F_m[n]\}_{m=0}^{N-1}$ називаються базисом Фур'є.

Перевіримо цей базис на ортогональність, для чого розрахуємо скалярний добуток двох довільних функцій з цього базису. Простір C^N є гільбертовим простором упорядкованих наборів з N чисел, скалярний добуток в ньому задається як:

$$\langle x, y \rangle = \sum_{k=0}^{N-1} x[k] \cdot y^*[k].$$

Тоді для функцій з базису (3.29) з номерами k та m скалярний добуток буде рівним:

$$\begin{aligned}\langle F_k[n], F_m[n] \rangle &= \sum_{n=0}^{N-1} F_k[n] \cdot F_m^*[n] = \\ &= \sum_{n=0}^{N-1} e^{2\pi j \frac{k}{N} n} \cdot e^{-2\pi j \frac{m}{N} n} = \begin{cases} 0, & (k-m) \neq 0, (k-m) \neq lN, l \in Z, \\ N, & (k-m) = 0, (k-m) = lN, l \in Z, \end{cases}\end{aligned}$$

отже, базис дискретних гармонічних функцій є ортогональним. Норма в гільбертовому просторі може бути задана як

$$\|x\|_2 = \sqrt{\langle x, x \rangle} = \left(\sum_{k=0}^{N-1} |x[k]|^2 \right)^{1/2}.$$

Норма базисної функції $F_m[n]$ буде рівна:

$$\|F_m[n]\| = \sqrt{\langle F_m[n], F_m[n] \rangle} = \sqrt{\sum_{n=0}^{N-1} e^{2\pi j \frac{m}{N} n} \cdot e^{-2\pi j \frac{m}{N} n}} = \sqrt{N}.$$

Часто використовують ортонормований дискретний базис Фур'є:

$$\begin{aligned}F_0[n] &= \frac{1}{\sqrt{N}} \\ F_1[n] &= \frac{1}{\sqrt{N}} e^{2\pi j \frac{n}{N}} \\ F_2[n] &= \frac{1}{\sqrt{N}} e^{4\pi j \frac{n}{N}} \\ F_3[n] &= \frac{1}{\sqrt{N}} e^{6\pi j \frac{n}{N}} \\ &\dots \\ F_{N-1}[n] &= \frac{1}{\sqrt{N}} e^{2\pi j (N-1) \frac{n}{N}}\end{aligned} \tag{3.30}$$

Коефіцієнти розкладу сигналу $x[n]$ в базисі Фур'є можна отримати за допомогою загального виразу для розрахунку коефіцієнтів розкладу сигналу в ортогональному базисі:

$$\begin{aligned}c[m] &= \frac{\langle x[n], F_m[n] \rangle}{\|F_m[n]\|^2} = \frac{\sum_{n=0}^{N-1} x[n] F_m^*[n]}{(\sqrt{N})^2}, \\ c[m] &= \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-2\pi j \frac{m}{N} n}.\end{aligned} \tag{3.31}$$

Маючи коефіцієнти розкладу сигналу $c[m]$, можна записати розклад сигналу $x[n]$ в базисі Фур'є:

$$x[n] = \sum_{m=0}^{N-1} c[m] e^{2\pi j \frac{m}{N} n}. \quad (3.32)$$

Вирази (3.31) та (3.32) називаються *прямим* та *оберненим дискретним перетворенням Фур'є дискретного скінченного сигналу*. В результаті прямого перетворення за відліками сигналу та значеннями базисних функцій можна отримати комплексні коефіцієнти розкладу $c[m]$ – спектр сигналу. За оберненим перетворенням, маючи коефіцієнти розкладу (комплексні амплітуди дискретних гармонік), можна отримати відліки сигналу $x[n]$.

Якщо використовувати ортонормований базис $\left\{ \frac{1}{\sqrt{N}} F_m[n] \right\}_{m=0}^{N-1}$ (3.30), то вирази для прямого і оберненого перетворення Фур'є дискретного сигналу матимуть такий вигляд:

$$c[m] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] e^{-2\pi j \frac{m}{N} n}, \quad (3.33)$$

$$x[n] = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} c[m] e^{2\pi j \frac{m}{N} n}. \quad (3.34)$$

Вирази (3.33)-(3.34) можна зустріти в літературі поряд з виразами (3.31)-(3.32). Обидві пари виразів для отримання спектру сигналу дають результат, який співпадає з точністю до масштабного множника. Це треба мати на увазі при інтерпретації результатів спектрального аналізу.

Отже, в результаті прямого перетворення Фур'є сигналу $x[n]$ тривалістю N відліків, ми отримаємо також N відліків комплексних спектральних складових $c[m]$, які будуть відповідати деяким частотам. Спектр можна представити в показниковій формі

$$c[m] = |c[m]| e^{j \arg(c[m])} = K[m] e^{j\Phi[m]},$$

де $K[m]$ – амплітудний спектр сигналу (АЧХ спектру $c[m]$);

$\Phi[m]$ – фазовий спектр дискретного сигналу (ФЧХ спектру $c[m]$).

АЧХ і ФЧХ є дискретними дійсними послідовностями, які складаються з N відліків. Оскільки сигнал $x[n]$ є дискретним, то його спектр $c[m]$ буде періодичним (наслідок явища періодизації спектру при дискретизації). Період буде дорівнювати N , що можна довести підстановкою:

$$c[N] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-2\pi j \frac{N}{N} n} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-2\pi j n} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{j0} = c[0]$$

$$c[N+1] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-2\pi j \frac{N+1}{N} n} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-2\pi j \frac{1}{N} n - 2\pi j n} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{2\pi j \frac{1}{N} n} = c[1]$$

$$c[N+2] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-2\pi j \frac{N+2}{N} n} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-2\pi j \frac{2}{N} n - 2\pi j n} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{2\pi j \frac{2}{N} n} = c[2]$$

...

Аналогічно, результат оберненого перетворення Фур'є (3.32) буде також періодичною послідовністю з періодом N :

$$x[N] = \sum_{m=0}^{N-1} c[m] e^{2\pi j \frac{N}{N} m} = \sum_{m=0}^{N-1} c[m] e^{-2\pi j m} = \sum_{m=0}^{N-1} c[m] e^{j0} = x[0]$$

$$x[N+1] = \sum_{m=0}^{N-1} c[m] e^{-2\pi j \frac{N+1}{N} m} = \sum_{m=0}^{N-1} c[m] e^{-2\pi j \frac{1}{N} m - 2\pi j m} = \sum_{m=0}^{N-1} c[m] e^{2\pi j \frac{1}{N} m} = x[1]$$

$$x[N+2] = \sum_{m=0}^{N-1} c[m] e^{-2\pi j \frac{N+2}{N} m} = \sum_{m=0}^{N-1} c[m] e^{-2\pi j \frac{2}{N} m - 2\pi j m} = \sum_{m=0}^{N-1} c[m] e^{2\pi j \frac{2}{N} m} = x[2]$$

...

Для номерів відліків від 0 до $N-1$ ця періодична послідовність буде співпадати з нашим початковим сигналом $x[n]$ (наш сигнал буде одним періодом деякої періодичної нескінченної послідовності). Через те, що з математичної точки зору обернене перетворення Фур'є дає нескінченну періодичну послідовність (в формулу (3.32) можна формально підставляти будь-які числа $m \in \mathbb{Z}$), ми маємо представлення начебто нескінченного періодичного сигналу. А отже, його спектр буде лінійчатим, заданим для кожного номеру m .

Якщо задатися питанням, чому дорівнює частота в герцах для відповідного спектрального відліку, ми не зможемо дати на це однозначну відповідь без додаткових відомостей про частоту дискретизації. Кожний відлік послідовності $c[m]$ є результатом скалярного добутку комплексної синусоїди (3.29) та сигналу $x[n]$. Точні значення частот цих синусоїд залежать від частоти дискретизації сигналу ω_s і від кількості відліків N .

Вводять поняття *основної частоти* при $m=1$, тобто $\omega_{0s} = \frac{\omega_s}{N} = \frac{2\pi}{NT_s}$, і всі частоти

гармонік будуть кратними цій основній частоті: $\omega[m] = m\omega_{0s} = m \frac{\omega_s}{N}$. Послідовності

N відліків сигналу, які виміряні з частотою дискретизації $\omega_s = \frac{2\pi}{T_s}$ (з кроком у часі T_s), відповідає спектр з N гармонічних складових, взятих з кроком $\frac{\omega_s}{N}$ (рис. 3.8).

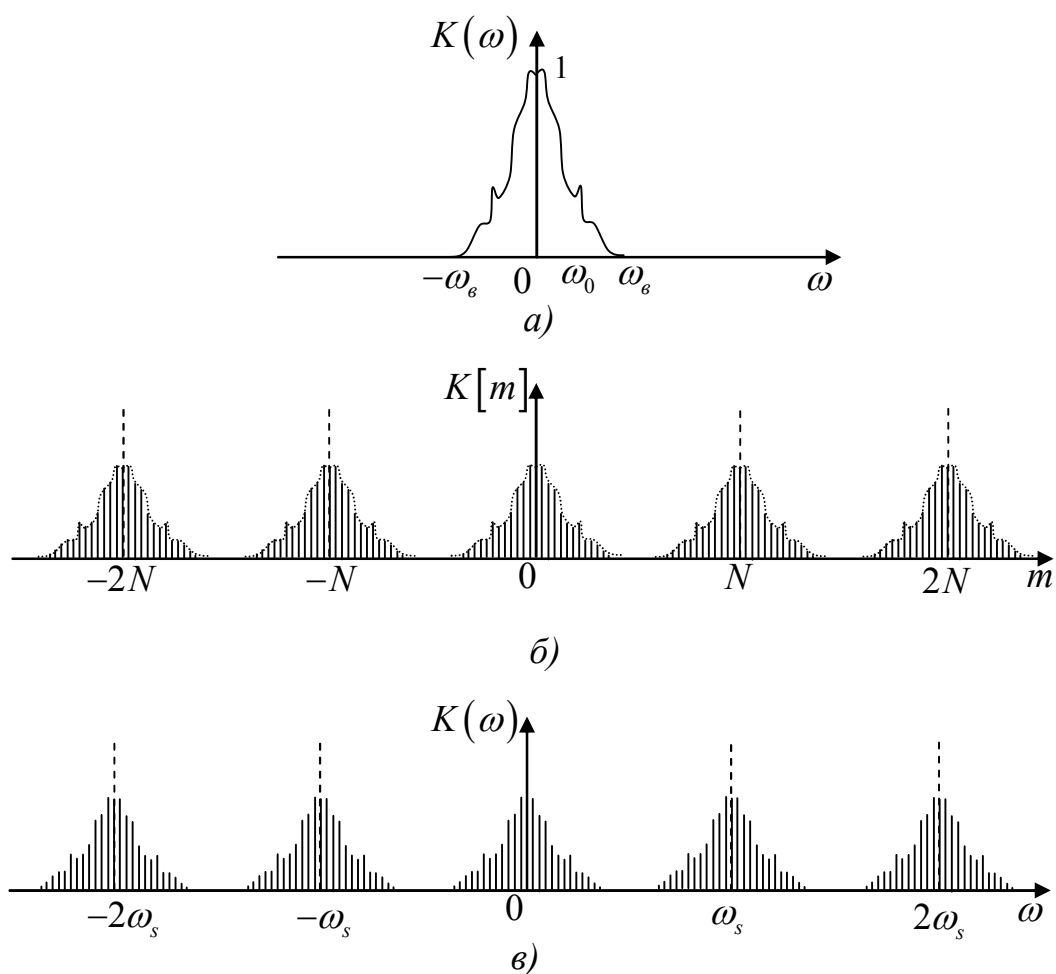


Рисунок 3.8 – а) амплітудний спектр початкового неперервного сигналу,

б) амплітудний спектр дискретного сигналу $x[n]$. По вісі частот відкладені номери відліків спектру m , відстань між спектральними лініями дорівнює 1

в) амплітудний спектр дискретного сигналу при відомій частоті дискретизації ω_s , відстань між спектральними лініями

дорівнює $\frac{\omega_s}{N}$ рад/с

Якщо в сигналі мало відліків, то значення $\frac{\omega_s}{N}$ буде великим, і різниця між частотами гармонічних складових в спектрі цього сигналу також буде великою. Чим більше відліків дискретного сигналу буде виміряно, тим меншою буде відстань $\frac{\omega_s}{N}$ між частотами сусідніх гармонік в спектрі цього сигналу. Іншими словами, чим більш тривалий сигнал піддається спектральному аналізу, тим більш «багатий» спектр ми отримаємо в результаті. На інтервалі частот між 0 та ω_s , де розраховується спектр дискретного сигналу, залежно від кількості відліків N буде розташовано більше чи менше частот, для яких будуть розраховуватися спектральні складові.

Наприклад, нехай частота дискретизації АЦП системи аналізу сигналів $F_s = 100$ Гц. Якщо буде зареєстровано сигнал тривалістю $T = 10$ секунд, то в нашому розпорядженні буде $N = 10 \times 100 = 1000$ відліків дискретного сигналу. Один період спектру цього сигналу буде розташований на частотах від 0 до 100 Гц, і відстань між гармоніками на частотній вісі буде рівною $\frac{F_s}{N} = \frac{100}{1000} = 0.1$ Гц. Тобто, в результаті

спектрального аналізу ми матимемо гармонічні складові на частотах 0, 0.1 Гц, 0.2 Гц, 0.3 Гц і т.д. аж до 100 Гц. Складових на будь-яких інших частотах в спектрі цього сигналу не буде, оскільки роздільна здатність визначається тривалістю сигналу. Якщо уявити, що нам потрібно розрізняти спектральні складові в сигналі, частоти яких відрізняються менше ніж на 0.1 Гц (наприклад, треба дослідити амплітуду складової з частотою 0.03 Гц або 57.23 Гц), то в даному випадку це неможливо, оскільки роздільна здатність надто мала.

Якщо збільшити час вимірювання сигналу до $T = 1000$ секунд, то матимемо вже $N = 1000 \times 100 = 100000$ відліків сигналу. Частоти гармонік в спектрі такого сигналу будуть відрізнятися на крок $\frac{F_s}{N} = \frac{100}{100000} = 0.001$ Гц. Роздільну здатність

спектрального аналізу стала більшою: тепер в спектрі можна розрізнити гармоніки, різниця частот яких є набагато меншою, тому буде відомо більше інформації про спектральні властивості сигналу, що аналізується.

Отже, тепер маємо інструментарій для спектрального аналізу дискретних сигналів, які задані на кінцевих проміжках часу, тобто тих сигналів, які переважно зустрічаються у реальному житті. Спектр за Фур'є таких сигналів є лінійчатим та періодичним, він повторюється з періодом, який дорівнює частоті дискретизації. Для практичних задач аналізу дійсних сигналів достатньо розрахувати значення спектру лише на одному періоді. Як правило, розрахунки проводять для значень $s[m]$ для номерів відліків від 0 до $N - 1$.

Збільшення тривалості сигналу приведе до підвищення роздільної здатності спектрального аналізу: можна буде розрізняти гармонічні складові, які мають близькі частоти.

3.10. Накладання спектрів

Той факт, що спектр дискретного сигналу є періодичною функцією частоти, має дуже важливий наслідок – можливість спотворення характеристик сигналу при його дискретизації.

Знаючи, що відбуватиметься зі спектром коли ми дискретизуємо сигнал, можна нарешті поставити питання: як саме треба дискретизувати. Умовою, яку треба виконати при проведенні дискретизації, є відсутність спотворень форми сигналу при перетворенні його з неперервного на дискретний. Іншими словами, аналіз дискретного сигналу має дати таку саме інформацію, ніби аналізується початковий неперервний сигнал.

Визначення умов, яким повинен задовольняти процес аналого-цифрового перетворення, проводиться в частотній області. Вимога неспотвореності форми сигналу тут формулюється як вимога неспотвореності спектру сигналу, тобто незмінності амплітуд його спектральних складових. Таке переформулювання можна зробити, оскільки перетворення Фур'є взаємно однозначно ставить у відповідність сигнал і спектр. Отже, необхідно проводити дискретизацію так, щоб форма спектру залишалася без змін. Якщо спектр не спотворено в результаті дискретизації, то теоретично можна відновити початковий неперервний сигнал, маючи лише його дискретні значення.

Розглянемо кілька випадків дискретизації одного і того самого сигналу з різними частотами дискретизації та подивимось, що може статися зі спектром. Нехай маємо сигнал $x(t)$ зі спектром як на рис. 3.9, а, який було дискретизовано з частотою ω_{1s} . Даний сигнал відноситься до класу сигналів з обмеженим спектром, тобто в його спектрі відсутні складові на частотах вищих ніж частота ω_c . Спектр результуючого дискретного сигналу буде періодичним (рис. 3.9, б).

Видно, що копії спектру сигналу (періоди або *репліки*), які відстоять на величину ω_{1s} один від одного, не перетинаються, не накладаються та не спотворюються. Кожен період спектру дискретного сигналу має таку саму форму, як і спектр неперервного сигналу. Якщо уявити, що дискретний сигнал пройде через систему, АЧХ якої $|H_f(j\omega)|$ має вигляд як вказано пунктиром на рис. 3.9, б, то на виході цієї системи буде сигнал, спектр якого лежатиме на частотах від $-\omega_c$ до ω_c . Тобто, на виході ми отримаємо неспотворений початковий неперервний сигнал зі спектром як на рисунку 3.9, а. Для цього система повинна мати межі смуги пропускання від $-\omega_c$ до ω_c . Такі системи-фільтри будуть розглянуті далі в цій книзі.

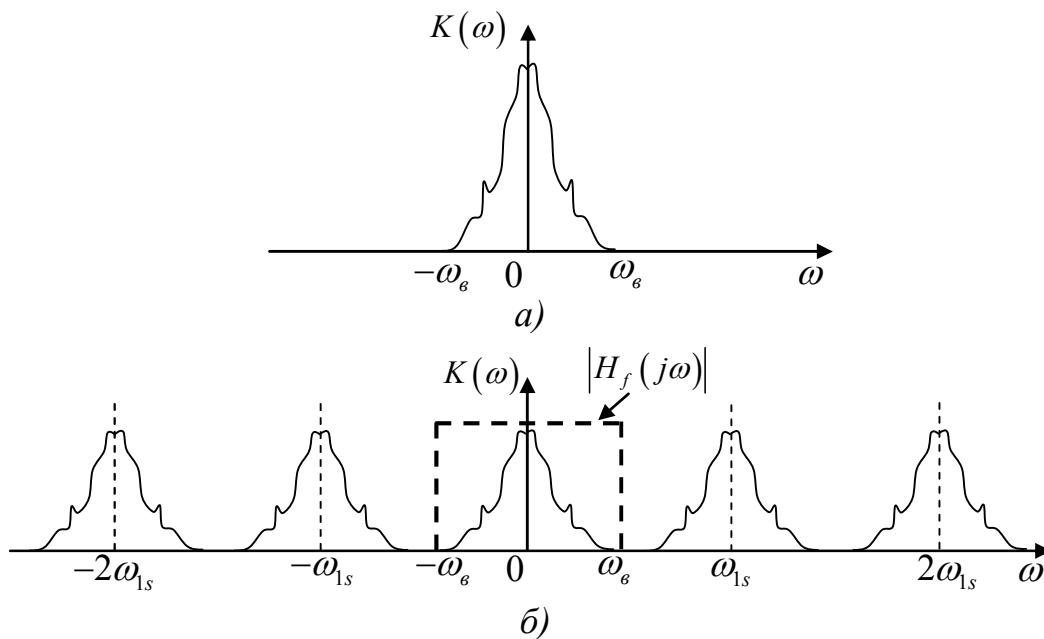


Рисунок 3.9 – а) амплітудний спектр початкового неперервного сигналу, б) амплітудний спектр того самого сигналу після дискретизації з частотою ω_{1s}

Отже, за умови, що період повторення спектру достатньо великий і спектр при дискретизації не спотворюється, всі властивості неперервного сигналу зберігаються в його дискретизованій версії. Неперервний сигнал може бути відновлений із дискретного.

Розглянемо інший випадок: нехай ми дискретизуємо той самий сигнал, але з частотою дискретизації $\omega_{2s} > \omega_{1s}$. Цей випадок подано на рис. 3.10, а. частота дискретизації стала більшою, отже більшим став період повторення спектру дискретного сигналу. Копії спектру початкового сигналу тепер відстоять одна від одної ще далі, вони не перетинаються і не накладаються. Отже, застосовуючи таку саме систему-фільтр з АЧХ $|H_f(j\omega)|$ як і в попередньому випадку, можна виділити з усього спектру дискретного сигналу лише один період, розташований в околі частоти $\omega = 0$, і відновити без втрат початковий неперервний сигнал.

Нехай тепер маємо АЦП, частота дискретизації якого менша ніж частота ω_{1s} : $\omega_{3s} < \omega_{1s}$. Спектр після дискретизації подано на рис. 3.10, б. Видно, що копії спектру тепер знаходяться ближче одна до одної, оскільки період повторення менший. Але ніяких спотворень поки що немає і неперервний сигнал можна відновити.

Сигнал не буде спотворений навіть у випадку, коли частота зрізу буде ще зменшена, то того значення ω_{4s} , коли сусідні копії спектру зрізних періодів доторкнуться одна до одної (рис. 3.10, в). Тим не менше, оскільки в межах від $-\omega_g$ до ω_g форма спектру не змінена відносно початкового спектру на рис. 3.9, а, то всі характеристики спектру збережені і сигнал може бути відновлений.

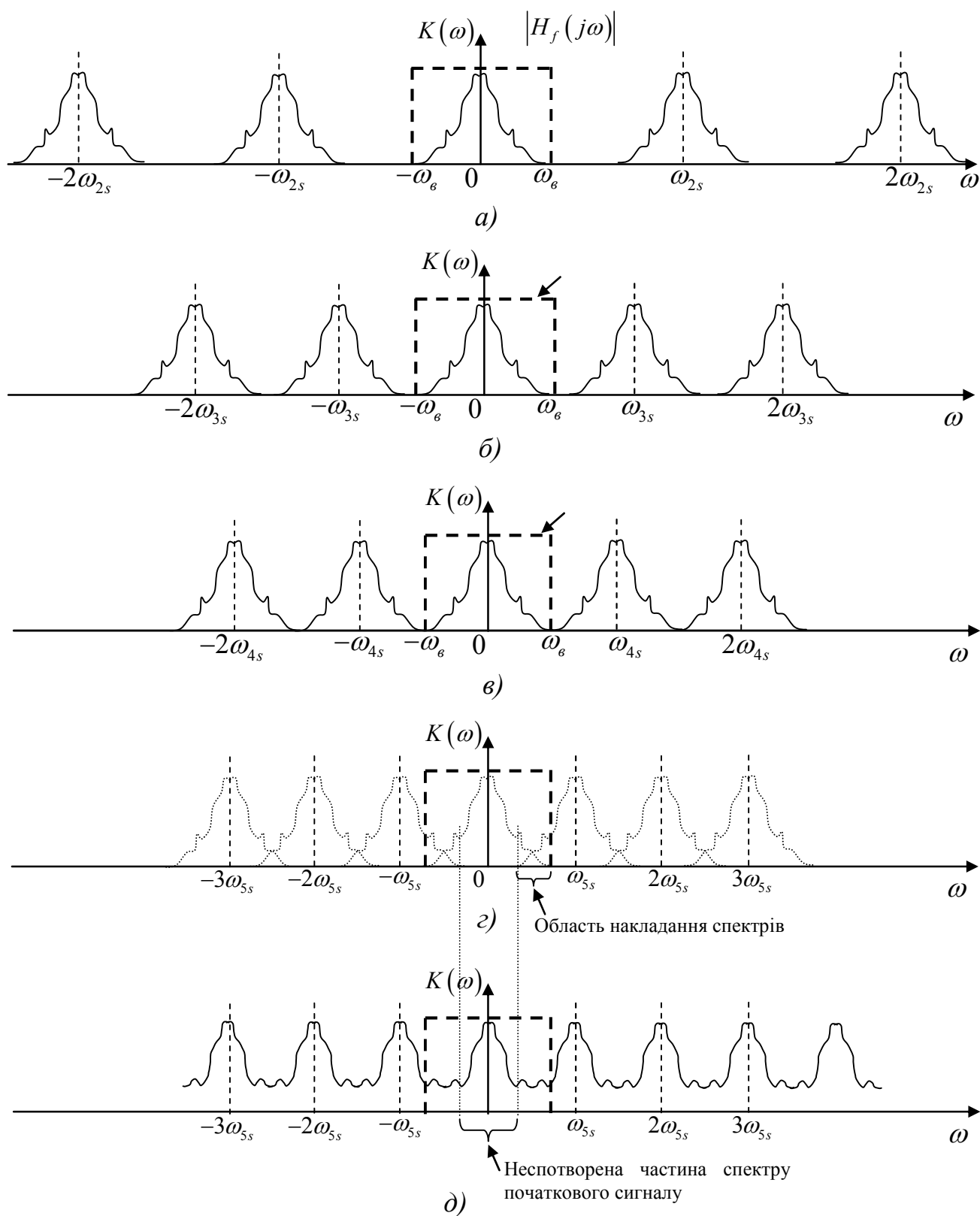


Рисунок 3.10 – Ілюстрація дискретизації сигналу з різними частотами дискретизації

Нехай ми ще зменшили частоту дискретизації, до значення ω_{s_s} (рис. 3.10, г). Вона вже настільки низька, що в межах одного періоду не поміщається весь спектр початкового сигналу. Видно, що копії спектру наче «накладаються» одна на одну, з'являється *область накладання спектрів* – область частот, на яких присутні одночасно спектральні складові, що відносяться до різних періодів періодичного спектру. Це веде до того, що на цих частотах в області перекриття збільшується амплітуда спектральних складових. Сусідні копії спектру накладаються одна на іншу, тобто ті частоти, які повинні бути у високочастотній частині спектру (ближче до ω_g), починають віддзеркалюватися, частини спектрів з різних періодів починають накладатися, внаслідок цього спотворюючи спектр дискретного сигналу. В результаті цього ефекту спектр сигналу, дискретизованого з занадто низькою частотою дискретизації, буде насправді мати вигляд, як на рис. 3.10, д. Частини цього спектру, які повинні були б відповідати періодам, буду сильно відрізнятися по формі від спектру початкового сигналу (рис. 3.9, а), їх навіть складно виділити з отриманого спектру.

Видно, що при накладанні спектрів вже неможливо виділити з періодичного спектру ту частину, яка точно відображає спектр початкового низькочастотного сигналу. Оскільки накладення відбувається починаючи з високих частот спектру, саме високочастотні складові і будуть спотворені в першу чергу. Неушкодженою залишиться лише частина спектру, яка знаходиться поза межами області накладання. Тому у випадку, коли неперервний сигнал був дискретизований з низькою частотою дискретизації, внаслідок накладання спектрів, дискретний сигнал вже більше не відображає всі властивості початкового неперервного сигналу. Якщо це не було враховано, то результати аналізу таких сигналів можуть бути хибними.

Описане явище називається *накладанням спектрів* (foldover, aliasing). Внаслідок такого явища неможливо після дискретизації точно відновити початковий неперервний сигнал по його відлікам. Відновлений сигнал буде спотвореним відносно початкового.

Розглянемо умови, при яких накладання спектрів не буде. Нехай маємо сигнал з обмеженим спектром від $-\omega_g$ до ω_g , дискретизований з частотою ω_s (рис. 3.11).

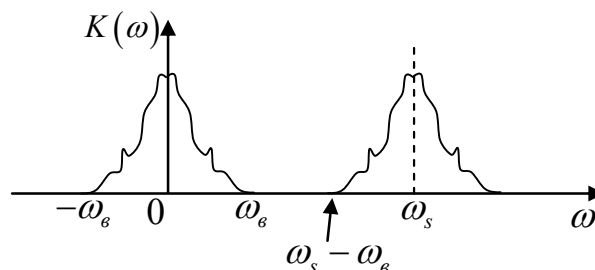


Рисунок 3.11 – До питання накладання спектрів

З ілюстрації видно, що у випадку, коли

$$\omega_s - \omega_g > \omega_g, \quad \omega_s > 2\omega_g \quad (3.35)$$

то копії спектрів, які знаходяться поряд, не накладаються, а отже, не спотворюються.

В цьому випадку ми можемо відновити початковий неперервний сигнал, застосувавши деяку спеціальну систему, яка виріже з всього періодичного спектру лише необхідну частину. Іншими словами, спектральні властивості початкового неперервного сигналу адекватно відображаються в отриманому дискретному сигналі.

3.11. Вибір частоти дискретизації. Теорема Котельникова-Шеннона

Ми можемо дискретизувати сигнал з будь-якою частотою, яка нам подобається, але виникає питання: наскільки добре представляють отримані значення початковий сигнал? В попередньому підрозділі було отримано правило, за яким можна обрати частоту дискретизації для сигналу, спектр якого лежить на частотах від $-\omega_c$ до ω_c . За цим правилом частота дискретизації повинна бути мінімум вдвічі більшою за верхню частоту спектру. Таке правило може бути сформульовано також виходячи з інших міркувань. Це альтернативне формулювання більше відоме в різних галузях науки і має багато назв. В класичній теорії сигналів воно називається теорема дискретизації або теорема відліків, в математиці – теорема Уїттакерів, в теорії інформації – теорема Шеннона, в теорії зв'язку – теорема Котельникова, в області передачі цифрових сигналів – теорема Найквіста. Першість у обґрунтуванні та доведенні теореми, яка дозволяє вибрати частоту дискретизації, належить В.О. Котельникову. Його теорема стала одним із теоретичних фундаментів всієї цифрової техніки і має величезне значення.

В 1932-33 рр. радянський інженер Володимир Олександрович Котельников задався питанням щодо того, чи можна передавати сигнал без спотворень по каналу зв'язку, смуга пропускання якого є меншою, ніж ширина спектра сигналу. Результатом його досліджень стала доповідь «О пропускной способности «эфира» и проволоки в электросвязи», яка була підготована для I-го Всесоюзного з'їзду з питань технічної реконструкції справи зв'язку та слабкострумної промисловості, який мав відбутися в 1933 році. Насправді, в його доповіді було доведено сім теорем, але всі вони були розвитком положень, що сформульовані в першій теоремі.

Теорема 1: «Будь-яку функцію $x(t)$, яка складається з частот від 0 до f_c , можна представити у вигляді ряду:

$$x(t) = \sum_{n=-\infty}^{+\infty} D_n \frac{\sin 2\pi f_c \left(t - n \frac{1}{2f_c} \right)}{t - n \frac{1}{2f_c}} \quad (3.36)$$

де D_n – деякі числа, що залежать від $x(t)$ ».

Функції під знаком суми називаються функціями Котельникова.

Теорема 2: «Будь-яку функцію $x(t)$, яка складається з частот від 0 до f_g можна неперервно передавати з будь-якою точністю за допомогою чисел, що слідують одне за одним через $\frac{1}{2f_g}$ секунд».

Оскільки в Теоремі 1 доведена рівність (3.36), то можна отримати значення $x(t)$ в моменти часу $t = n\frac{1}{2f_g}$, які запропоновані в Теоремі 2. Підставимо їх в вираз, та розкриємо невизначеність виду $\frac{\sin x}{x}$:

$$\begin{aligned} x\left(n\frac{1}{2f_g}\right) &= \sum_{n=-\infty}^{+\infty} D_n \frac{\sin 2\pi f_g \left(n\frac{1}{2f_g} - n\frac{1}{2f_g}\right)}{n\frac{1}{2f_g} - n\frac{1}{2f_g}} = \\ &= \sum_{n=-\infty}^{+\infty} D_n \frac{\sin 2\pi f_g \left(n\frac{1}{2f_g} - n\frac{1}{2f_g}\right)}{\frac{2\pi f_g}{2\pi f_g} \left(n\frac{1}{2f_g} - n\frac{1}{2f_g}\right)} = 2\pi f_g D_n. \end{aligned}$$

Розглянемо вираз (3.36). При відомій верхній граничній частоті спектру f_g , для того, щоб знати точний вираз для неперервного сигналу $x(t)$, достатньо лише знати числа D_n . З попереднього виразу видно, що якщо ми будемо в кожний момент часу $t = n\frac{1}{2f_g}$ вимірювати значення $x\left(n\frac{1}{2f_g}\right)$, та розраховувати значення

$$D_n = \frac{x\left(n\frac{1}{2f_g}\right)}{2\pi f_g},$$

то можна точно записати вираз для неперервного сигналу $x(t)$ (3.36). Якщо ці значення D_n , або лише значення $x\left(n\frac{1}{2f_g}\right)$, будуть передаватися по деякому каналу зв'язку, то на стороні приймача можна буде точно відновити неперервний сигнал $x(t)$ на основі цих чисел.

Отже, завдяки властивостям, доведеним в Теоремах 1 та 2, існує можливість замінити неперервний сигнал з обмеженим спектром, його дискретними відліками, які взяті з частотою дискретизації $2f_g$.

Ця вимога для вибору частоти дискретизації сигналу з обмеженим спектром від 0 до f_g часто називається «теоремою Котельникова», що не зовсім коректно. Насправді, вона випливає з двох теорем, доведених Котельниковим, і є їхнім наслідком. Теореми Котельникова охоплюють набагато ширше коло питань, ніж суто вибір частоти дискретизації сигналу.

Доцільно також навести для порівняння теорему Шеннона:

«Якщо спектр сигналу $x(t)$ не містить частот вище за f_g Гц, то він повністю визначається своїми миттєвими значеннями в моменти, що віддалені один від одного на $\frac{1}{2f_g}$ секунд».

Є також загальноприйняте формулювання, яке хибно називають «теоремою Котельникова»: «Частота дискретизації сигналу, спектр якого лежить від 0 до f_g Гц, повинна бути не меншою, ніж $2f_g$ Гц». Це правило говорить про нижню межу значень частоти дискретизації, в реальних системах частота дискретизації часто обирається набагато більшою.

Отже, ми отримали *правило, за яким можна розрахувати частоту дискретизації* для сигналів, які ми аналізуємо: для того, щоб всі властивості неперервного сигналу, які відображені в його спектрі за Фур'є, збереглися в дискретизованому сигналі, частота дискретизації повинна бути більша за подвоєну максимальну частоту ненульової гармоніки в спектрі сигналу.

Як видно з рисунків до попереднього підрозділу, частота дискретизації, обрана за йим правилом – це мінімальна частота, з якою теоретично можливо проводити дискретизацію, щоб в спектрі не було спотворень внаслідок накладання.

Максимальна теоретично можлива частота, що може бути присутньою в спектрі сигналу, називається *частотою Найквіста*. Вона дорівнює половині частоти дискретизації. Це та частота, в районі якої відбувається накладення спектрів (частота завороту, частота перегину).

Дискретизація з такими вимогами до частоти дискретизації часто називається *низькочастотною дискретизацією*, оскільки спектр сигналу, який дискретизують, зосереджений на нижніх частотах в околі нульової частоти.

Зауваження до вибору частоти дискретизації.

1. Безпосереднє використання вибору частоти дискретизації за Котельниковим можливе за умови, коли дослідник впевнений, що спектр сигналу є обмеженим, і за межами верхньої частоти спектральних складових немає. На практиці це може означати, що сигнал попередньо відфільтрований, щоб видалити складові на вищих частотах, або що система обробки сигналів не чутлива до спектральних складових на вищих частотах.

2. На практиці для позбавлення від ефекту накладення спектрів необхідно використовувати попередню фільтрацію за допомогою фільтру нижніх частот. Цей фільтр включений до стандартної структури системи цифрової обробки сигналів, яка розглядалася в розділі 2.

На рис. 3.12, а, зображено спектр деякого неперервного сигналу з достатньо широким спектром. На вісі частот показано, яку частоту дискретизації має АЦП. Видно, що при дискретизації з такою частотою буде відбуватися накладення спектрів. Щоб накладення спектрів не було, необхідно, щоб в спектрі сигналу були відсутні складові на частотах від $-\omega_N$ до ω_N , де ω_N – частота Найквіста (половина частоти дискретизації). Отже, перед тим, як дискретизувати сигнал з частотою ω_s , необхідно позбутися від спектральних складових в межах від ω_N до $+\infty$. Це можна зробити, пропустивши неперервний сигнал через фільтр, який має АЧХ $|H_f(j\omega)|$, показану пунктиром. Коефіцієнт передачі в межах від $-\omega_N$ до ω_N повинен дорівнювати одиниці, а поза цим проміжком – нулю.

На рис. 3.12, б, подано спектр результуючого сигналу, в якому немає складових на частотах, вищих за половину частоти дискретизації. Для такого сигналу виконуються умови теореми дискретизації, тому його можна дискретизувати з частотою $\omega_s = 2\omega_g$. На рис. 3.12, в, показано, що накладення спектрів в такому випадку відсутнє, бо немає частотних складових, які будуть накладатися в сусідніх періодах. Остаточний спектр дискретного сигналу, отриманого в результаті дискретизації, наведено на рис. 3.12, г. Видно, що існує можливість відновити з такого дискретного сигналу неперервний за допомогою фільтра. Отриманий неперервний сигнал буде мати спектр як на рис. 3.12, б. Треба зазначити, що початкового неперервного сигналу зі спектром 3.12, а, отримати вже не вдасться, оскільки операція фільтрації є незворотною.

В результаті таких операцій було виключене явище накладення спектрів, але натомість втрачено ті складові спектру сигналу, які лежали вище частоти Найквіста. Імовірно, що це не оптимальний спосіб дослідження спектральних властивостей такого сигналу, оскільки частотні складові на частотах, які було вирізано, могли нести корисну інформацію про сигнал. Правильно було б взяти інший АЦП з більш високою частотою дискретизації.

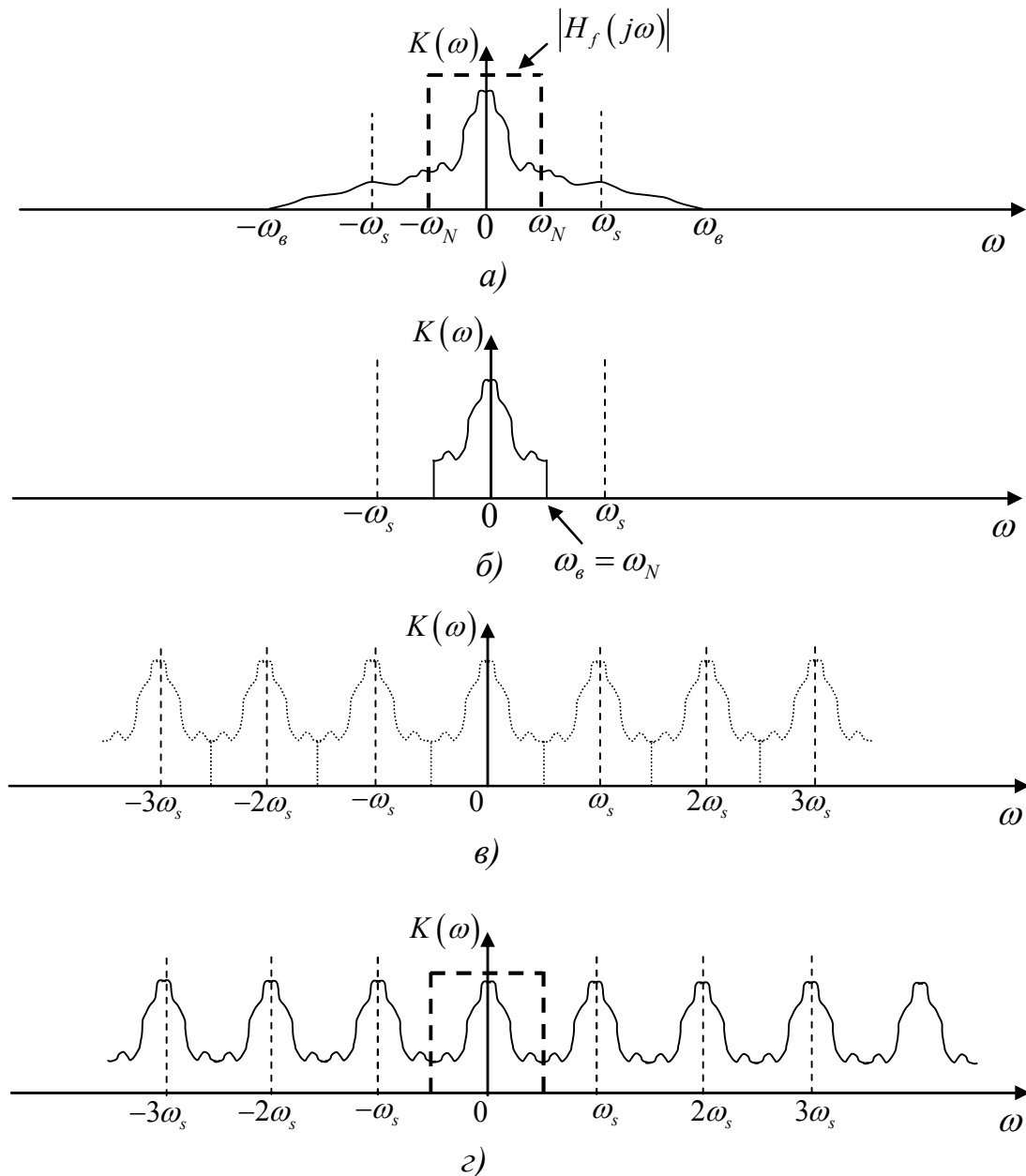


Рисунок 3.12 – Попередня фільтрація неперервного сигналу для попередження накладання спектрів

3. При обробці реальних сигналів дуже рідко постає задача подальшого відновлення неперервного сигналу по його дискретним відлікам. Частіше ставиться задача лише отримати спектр досліджуваного сигналу та проаналізувати його. При цьому намагаються дискретизувати з якомога більшою частотою дискретизації.

4. При безпосередньому використанні теореми Котельникова щодо відновлення неперервного сигналу за послідовністю чисел, сигнал має бути відновленим за (3.36)

. Цей вираз є сумою нескінченної кількості функцій вигляду $\frac{\sin x}{x}$, які зміщені по

часовій вісі на цілу кількість кроків дискретизації $\frac{1}{2f_s}$, та кожна з яких має свою величину D_n (рис. 3.13, показані пунктиром). Щоб (3.36) справджувалася, необхідно, щоб ці функції були визначені на проміжку від $-\infty$ до $+\infty$.

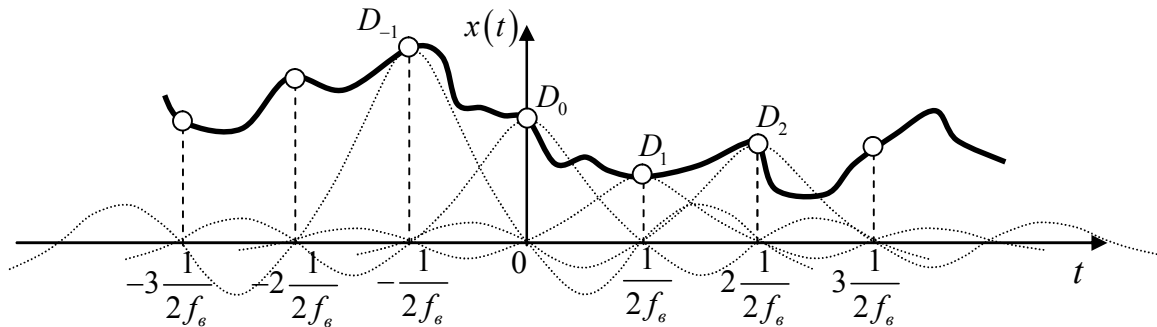


Рисунок 3.13 – Відновлення неперервного сигналу за Котельниковим

Але проблема в тому, що для того, щоб згенерувати такі функції, для кожного моменту часу (відліку) $n\frac{1}{2f_s}$, необхідно знати їх величину D_n . Ця величина

однозначно пов'язана зі значенням сигналу, який ми дискретизуємо, в момент відліку. Отже, для того, щоб в даний момент часу мати функцію, необхідно було починати її генерувати з часу, що дорівнює $-\infty$. І заздалегідь знати, чому буде дорівнювати її величина зараз. Так само треба зробити для всіх функцій. Іншими словами, для того, щоб відновити сигнал за допомогою функцій Котельникова за (3.36), треба заздалегідь знати, якими будуть всі відліки сигналу. Авжеж, майбутні значення відліків сигналу прогнозувати неможливо. Отже, хоча теоретично (3.36) є вірною, на практиці нею скористатися не можна. Не існує жодної фізичної системи, яка здатна відновити нескінченний неперервний сигнал за Котельниковим.

3.12. Дискретизація смугових сигналів

В попередньому підрозділі було розглянуто ситуацію, коли спектр неперервного сигналу, що підлягає дискретизації, сконцентрований в низькочастотній частині спектру навколо частоти $\omega=0$. В цьому випадку стандартним вибором по визначенню частоти дискретизації є $\omega_s \geq 2\omega_g$, де ω_g –максимальна частота ненульової гармонічної складової спектру.

Розглянемо іншу ситуацію, коли спектр сигналу, який треба дискретизувати, займає смугу частот, яка починається не від нульової частоти, а від деякої частоти ω_L до ω_H (рис. 3.14). Такі сигнали широко використовуються в сучасних системах

мобільного зв'язку системах RFID, WiMAX та т.і., дані в яких передаються за допомогою вузькосмугових високочастотних сигналів.

Вузькосмуговим сигналом називається сигнал, ширина смуги якого значно менша його центральної частоти:

$$\frac{\omega_H - \omega_L}{\omega_c} = \frac{\Delta\omega}{\omega_c} \ll 1, \quad \omega_c = \frac{\omega_L + \omega_H}{2}.$$

Це сигнали в системах цифрового радіо, мобільного зв'язку з частотним розділенням каналів, оптиці, радарів, локальних безпроводних комп'ютерних мережах та ін.

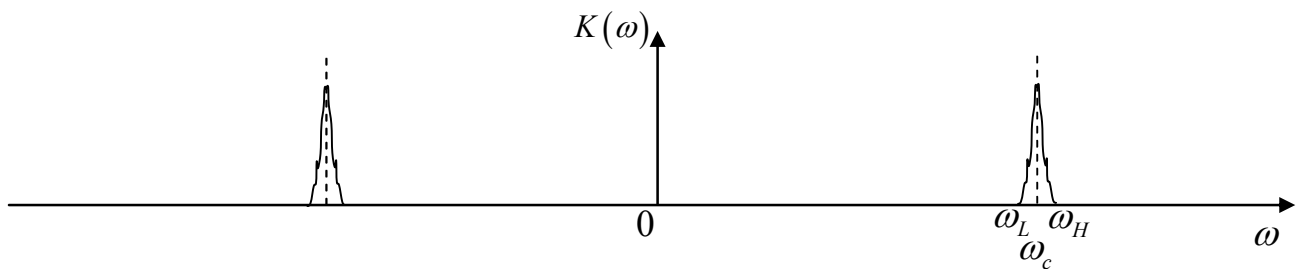


Рисунок 3.14 – Спектр вузькосмугового сигналу, який займає смугу частот на високих частотах

Ширина смуги частот, яку займає сигнал, часто дуже мала порівняно з верхньою та нижньою граничними частотами в спектрі, тому використовувати підхід низькочастотної дискретизації за Котельниковим неекономно. Частота дискретизації, обрана безпосередньо з умови $\omega_s \geq 2\omega_H$ буде надлишково високою, особливо для радіосигналів, де вона може скласти сотні мегагерц і навіть гігагерци. Отже, обробка в реальному часі може бути складною або неможливою, або внаслідок складності алгоритмів, або в наслідок обмеженості ресурсів існуючої елементної бази.

Для дискретизації таких сигналів можна використати три методи:

- метод «смугової дискретизації» (синоніми: субдискретизації, undersampling, bandpass sampling, subsampling, super-Nyquist, sub-Nyquist, harmonic sampling);
- дискретизація на основі переходу до аналітичного сигналу (аналітичний сигнал – це комплекснозначний сигнал, спектр якого при додатних частотах співпадає зі спектром нашого сигналу, а для від'ємних частот дорівнює нулю);
- дискретизація на основі виділення квадратурних складових (виконується шляхом зсуву спектру досліджуваного сигналу в область низьких частот з використанням гетеродинів – генераторів проміжних частот).

Всі ці методи дозволяють знизити потрібну швидкодію АЦП та зменшити об'єм пам'яті, потрібний на збереження сигналу.

Розглянемо більш докладно метод смугової дискретизації.

Згадаємо, що при дискретизації неперервного сигналу спектр дискретного сигналу стає періодичним, а період повторення спектру дорівнює частоті дискретизації.

Дивлячись на нескінченний періодичний спектр дискретизованого сигналу, неможливо сказати, якою була частота дискретизації (відомо лише, що вона повинна бути кратною періоду спектру), а також неможливо сказати, який саме з періодів спектру відповідає спектру початого неперервного сигналу (оскільки вони мають однаковий вигляд) (рис. 3.15).

Можна сказати, що яким би не був спектр сигналу, і де б на вісі частот він не був розташований, після дискретизації спектр стане періодичним. Спектральні складові початкового неперервного сигналу, які можуть бути розташовані будь-де на вісі частот, будуть повторюватися періодично на всій вісі.

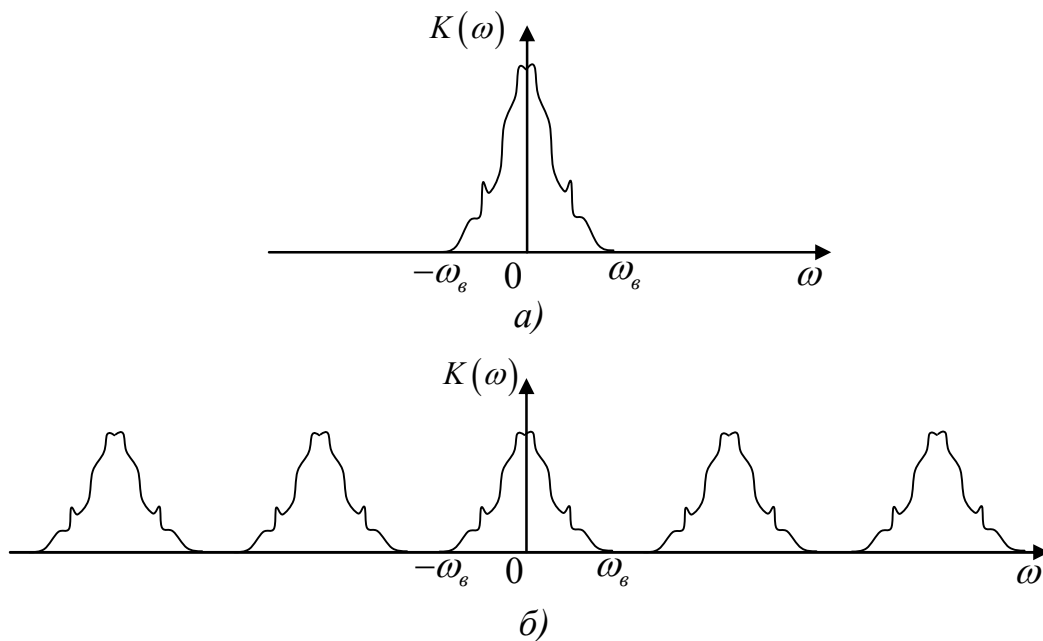


Рисунок 3.15 – а) амплітудний спектр початкового неперервного сигналу, б) амплітудний спектр того самого сигналу після дискретизації

Ця властивість можна використати для того, щоб суттєво знизити необхідну частоту дискретизації. Нехай маємо неперервний сигнал зі спектром як на рисунку 3.14, який дискретизуємо з частотою ω_s . Після дискретизації спектр стане періодичним і простягнеться по всій вісі частот (рис. 3.16). Між частотами ω_L та ω_H спектр дискретного сигналу буде співпадати зі спектром початкового неперервного сигналу.

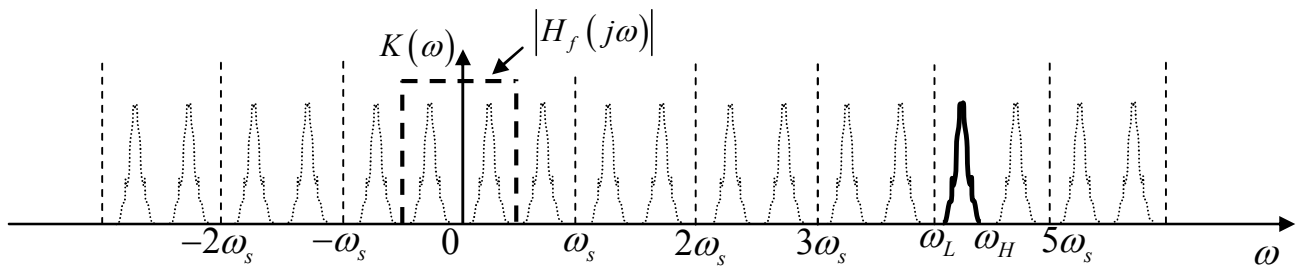


Рисунок 3.16 – Спектр дискретизованого вузькосмугового сигналу, який займає смугу частот на високих частотах

Видно, що в результаті дискретизації копії спектру будуть не тільки на високих частотах, а і на частотах в околі нульової частоти. Всі копії ідентичні, і форма спектру для всіх них однакова. Отже, для того, щоб отримати інформацію про властивості спектральної характеристики сигналу, можна використати будь-яку з отриманих копій. А для того, щоб мати уяву про точні значення частот всіх гармонічних складових, необхідно знати, в якому саме періоді знаходиться спектр аналізованого сигналу, що можливо на практиці. Видно, що для відновлення неперервного сигналу з його дискретного представлення достатньо скористатися фільтром нижніх частот, АЧХ якого представлена на рисунку. На виході цього фільтру буде низькочастотний вузькосмуговий сигнал, форма спектру якого така сама, як у початкового неперервного сигналу.

Як і завжди, дискретизація з частотою, набагато нижчою від подвоєної верхньої граничної частоти спектру сигналу, повинна відбуватися за умови, що накладення копій спектрів на нижніх частотах відсутнє. Отримаємо умови вибору частоти дискретизації, за яких це буде мати місце.

Розглянемо той інтервал частот, на якому розташований спектр нашого початкового неперервного сигналу (рис. 3.17) при дискретизації з частотою ω_s . Видно, що перекриття копій спектру та спотворення спектру внаслідок явища накладення буде відсутнім тоді, коли спектр нашого неперервного сигналу розташований тільки в межах одного півперіоду майбутнього періодичного спектру. Довжина періоду залежатиме від того, яку ми оберемо частоту дискретизації. Якщо частота дискретизації підібрана таким чином, що ця умова дотримана, то після дискретизації накладення спектрів не буде на всій частотній вісі, а, отже, і в околі нижніх частот. На основі цього міркування можна вибрати частоту дискретизації вузькосмугового високочастотного сигналу.

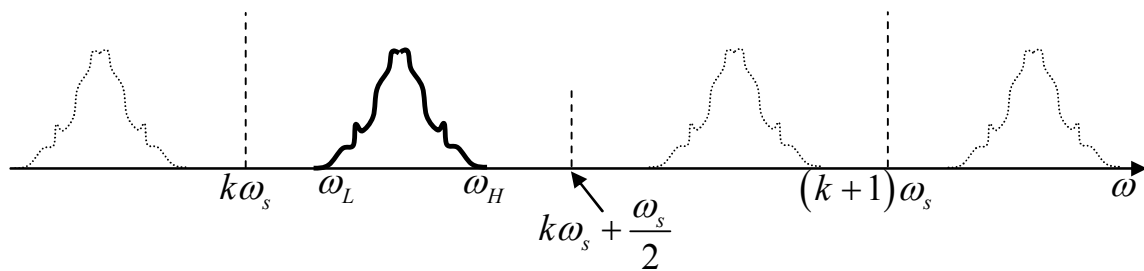


Рисунок 3.17 – Визначення умов відсутності накладання спектрів при дискретизації смугових сигналів

Нехай спектр початкового сигналу опинився на тому періоді, який лежатиме між частотами $k\omega_s$ та $(k+1)\omega_s$. Серединою цього періоду в спектрі буде деяка частота $k\omega_s + \frac{\omega_s}{2}$. Для випадку, зображеного на рисунку 3.17, спектр сигналу повинен

знаходитись між частотами $k\omega_s$ та $k\omega_s + \frac{\omega_s}{2}$, тобто мають виконуватися такі умови:

$$\frac{\omega_s}{2} > \omega_H - \omega_L,$$

$$k\omega_s < \omega_L,$$

$$\omega_H < k\omega_s + \frac{\omega_s}{2}.$$

Переписавши нерівності для частоти дискретизації, отримаємо

$$\omega_s \geq 2\Delta\omega,$$

$$\frac{2\omega_H}{2k+1} < \omega_s < \frac{\omega_L}{k}.$$

тут k – номер діапазону частот, який повинен приймати тільки цілі значення. Перегрупувавши сторони останньої нерівності та позбувшись частоти дискретизації, отримаємо:

$$k < \frac{\omega_L}{2\Delta\omega},$$

що є границею зверху для номеру діапазону. Вибір конкретного діапазону залежить від наявних технічних можливостей (АЦП), можливості реалізації фільтрів, спектральних характеристик шумів та інших вимог.

Отже, можна відновити неперервний високочастотний вузькосмуговий сигнал по копіям його спектру; для цього треба обрати частоту дискретизації так, щоб не відбувалось накладання спектрів.

Мінімальна теоретично можлива частота дискретизації при цьому визначається згідно із ще одною теоремою, доведеною Котельниковим:

«Будь-яку функцію $x(t)$, яка складається з частот від f_1 до f_2 , можна неперервно передавати з будь-якою точністю за допомогою чисел, що слідують одне за одним через $\frac{1}{2(f_2 - f_1)}$ секунд.»

Наслідком з цієї теореми є вимога до частоти дискретизації: для смугового сигналу вона має бути не меншою, ніж подвоєна ширина спектру сигналу: $F_s \geq 2(f_2 - f_1)$.

Котельников також довів, що будь-яка функція, що складається з частот від f_1 до f_2 , може бути представлена у вигляді:

$$f(t) = F_1(t) \cos \frac{2\pi f_1 + 2\pi f_2}{2} t + F_2(t) \sin \frac{2\pi f_1 + 2\pi f_2}{2} t,$$

де $F_1(t)$ та $F_2(t)$ – деякі функції, що складаються з частот від 0 до $\frac{f_2 - f_1}{2}$.

Для відновлення сигналу можна використати фільтр нижніх частот, який виріже зі спектру лише потрібну смугу.

При смуговій дискретизації реальних сигналів треба не забувати про те, що в сигналі присутня також і шумова складова. Як правило, спектр корисного сигналу відомий досліднику заздалегідь, але спектральні характеристики шумів апіорі точно невідомі. Отже, коли будуть дотримані всі умови для відсутності накладення спектру корисного сигналу, шумова складова може піддаватися накладанню, і зокрема може опинитися в тій області частот, де буде знаходитись спектр корисного сигналу. В цьому випадку спектр корисної складової буде спотворений внаслідок накладення спектрів. В загальному випадку, всі складові, які знаходяться в спектрі сигналу поза межами смуги від ω_L до ω_H , будуть піддаватися накладенню. Тому необхідно враховувати цю можливість та позбавлятися від шумів заздалегідь.

3.13. Ефект Гіббса

Розглянемо один з ефектів, який виникає при обробці реальних сигналів та який треба враховувати при аналізі спектрів. Розглянемо неперіодичний неперервний сигнал $x(t)$ (Рис. 3.18, а), властивості якого треба дослідити. Нехай він має спектр, як зображено на рисунку 3.18, б.

Коли ми проводимо вимірювання, сигнал задається лише на певному проміжку часу, який ми самі обмежуємо та обираємо для подальшої обробки. Наприклад, нехай вимірювання значень сигналу припинилося в момент часу T_1 . Отже, для подальшого аналізу замість всього неперервного нескінченного сигналу ми маємо в своєму розпорядженні лише його частину для проміжку часу від 0 до T_1 . Для проведення спектрального аналізу ми маємо розрахувати представлення цього сигналу у вигляді суми гармонічних складових. Ці складові разом повинні дати наш сигнал:

$$x(t) = \sum_{n=0}^{+\infty} A_n \cos(n\omega_0 t + \varphi_n).$$

Видно, що наш сигнал складатиметься з періодичних нескінченних косинусоїдальних сигналів, тому і сума таких гармонік буде періодичним сигналом. З точки зору математики, ділянка сигналу, яка підлягає аналізу за Фур'є, та представлена у вигляді розкладу на гармоніки, буде періодично повторюватися на всій часовій вісі. Іншими словами, якщо взяти значення за спектру амплітуди та початкові фази всіх гармонік, згенерувати такі гармоніки і скласти за гармонічних складових сигнал, то він буде періодичним і нескінченним. Інтервал від 0 до T_1 буде його періодом, і на цьому періоді отриманий сигнал буде співпадати з початковим сигналом $x(t)$. Отже, спектр сигналу, який буде отримано в результаті перетворення Фур'є, представлятиме початковий сигнал на проміжку часу від 0 до T_1 , а також періодичне повторення початкового сигналу поза межами цього проміжку (рис. 3.18, в).

Розглянемо тепер властивості цього нескінченного періодичного сигналу, який описується отриманим спектром, та на його зв'язок з початковим нескінченним сигналом $x(t)$, який ми хотіли аналізувати. Зрозуміло, що чим більший проміжок сигналу піддається аналізу, тим більше властивості всього нескінченного сигналу будуть відображені в ділянці від 0 до T_1 , і тим більше спектр ділянки сигналу буде схожий на спектр самого сигналу.

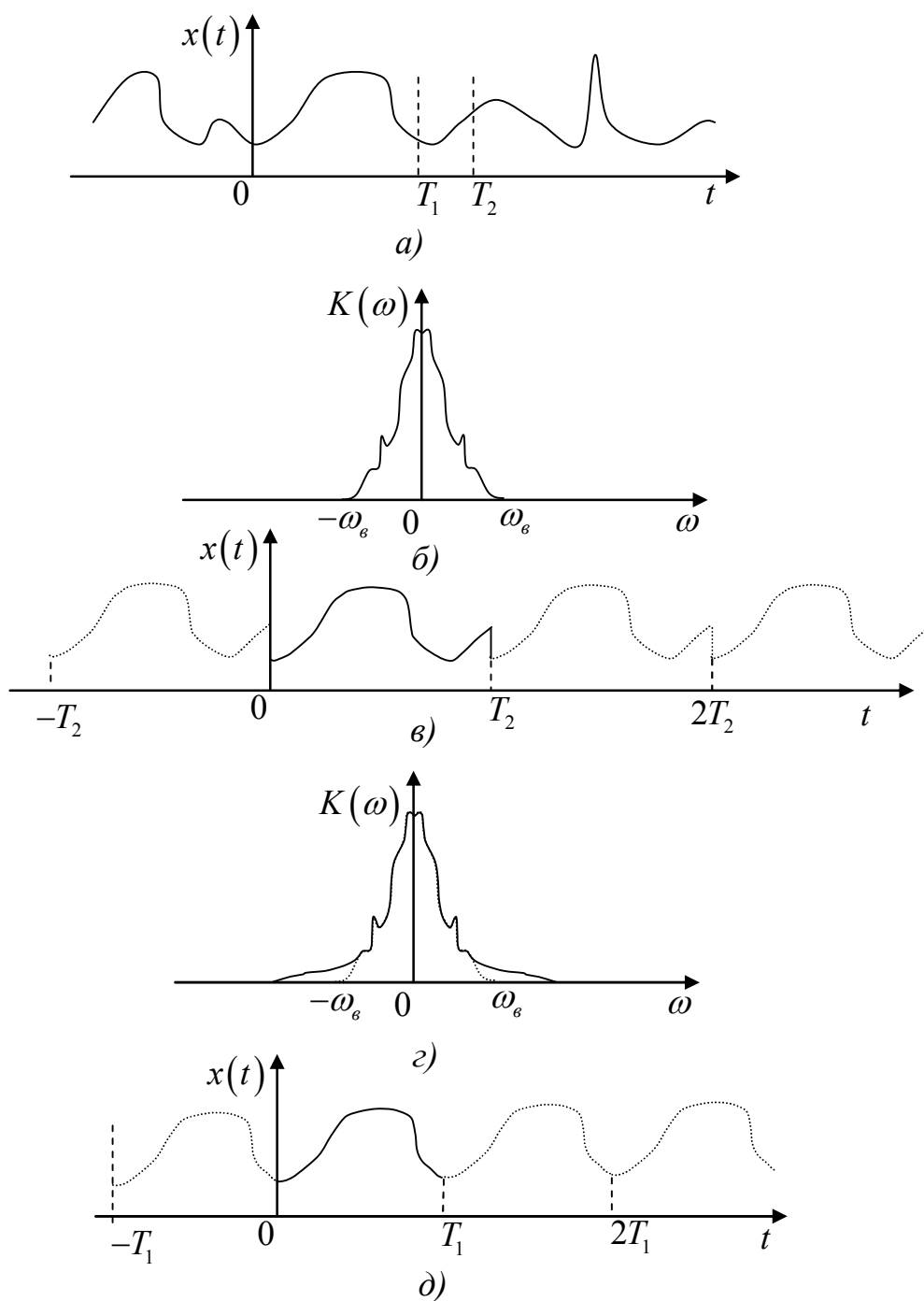


Рисунок 3.18 – а) досліджуваний нескінченний неперервний сигнал, б) його амплітудний спектр, в) періодичний сигнал, нерівні значення на початку і в кінці проміжку, г) спектр сигналу з рис. в), д) періодичний сигнал, значення на початку і в кінці проміжку співпадають

Поглянемо також на початок і кінець інтервалу вимірювання. Можливі дві ситуації: значення сигналу на початку і в кінці проміжку або співпадають, або не співпадають. Якщо буде ситуація, коли значення сигналу на початку та в кінці проміжку не рівні одне одному, при періодичному повторенні виміряної ділянки в

періодичному сигналі будуть присутні розриви першого роду (скінченні стрибки) (Рис. 3.18, в). Ці розриви є швидкими змінами значення сигналу, які повинні відображатися спектральними складовими високих частот. Для такого сигналу ці стрибки будуть обов'язково відображені у спектрі. В спектральній області це приведе до зростання амплітуд гармонічних складових на верхніх частотах (рис. 3.18, г). В початковому нескінченному сигналі таких високочастотних складових немає, вони з'явилися в спектрі лише внаслідок невдалого вибору проміжку, на якому вимірюється сигнал. Це явище називається ефектом Гіббса.

У випадку, коли значення на початку і в кінці інтервалу рівні одне одному (рис. 3.18, д), додаткові спектральні складові в спектрі не з'являються і спектр не спотворюється.

3.14. Ефект розтікання спектру

Розглянемо ще один ефект, який треба знати для більшого розуміння сутності спектрального аналізу за Фур'є та його практичного застосування. Нехай маємо скінченний дискретний сигнал, що складається з N відліків, який отриманий з частотою дискретизації ω_s . Поставимо собі задачу отримати спектр цього сигналу за Фур'є. З попередніх розділів відомо, що для цього треба сформувати набір дискретних базисних функцій Фур'є в N -вимірному просторі, і для отримання коефіцієнтів розкладу (спектру) необхідно розрахувати скалярні добутки аналізованого сигналу на базисні функції за виразом:

$$c_x[m] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-2\pi j \frac{m}{N} n}.$$

Розглянемо частоти, на яких буде отримано складові спектру даного сигналу. Сигнал, що складається з N відліків, є вектором в N -вимірному просторі. Базис такого простору складається з N функцій, тому в результаті розрахунків буде визначено стільки ж коефіцієнтів розкладу (проекцій сигналу на базисні функції), скільки базисних функцій було залучено, а саме N . Відомо, що спектр дискретного сигналу є періодичним, і період дорівнює частоті дискретизації. На кожному періоді спектру буде отримано N відліків, які відповідають частотам, рівномірно розташованим між 0 та ω_s . Інтервал між сусідніми частотними складовими дорівнює

основній частоті $\omega_{0s} = \frac{\omega_s}{N} = \frac{2\pi}{NT_s}$. Видно, що чисельні значення частот гармонічних

складових, на які буде розкладено сигнал, залежать від частоти дискретизації ω_s та від тривалості сигналу N . Від тривалості сигналу також залежатиме кількість коефіцієнтів розкладу.

Отже, розклад сигналу за Фур'є

$$x[n] = \sum_{m=0}^{N-1} c[m] e^{2\pi j \frac{m}{N} n}$$

буде вестися на гармонічні складові, частоти яких напряду зв'язані з частотою дискретизації та тривалістю сигналу. Частоти кожної гармонічної функції в сформованому базисі дорівнюють $\omega[m] = m \frac{\omega_s}{N}$, де $m = \overline{0, N-1}$ – номер гармоніки.

При розкладі сигналу ми будемо мати коефіцієнти розкладу лише для таких гармонік.

Поглянемо на ситуацію з іншого боку. Властивості сигналу, який піддається аналізу, заздалегідь невідомі; власне, для того і виконують спектральний аналіз за Фур'є, щоб дослідити спектр сигналу і зробити певні висновки. Тому апріорі невідомо, які коливальні складові будуть в складі даного сигналу: ми не знаємо до аналізу, з *яких гармонік* буде складений сигнал, що ми досліджуємо. Через це неможливо заздалегідь так підібрати тривалість сигналу та частоту дискретизації, щоб у базисі для розкладу даного сигналу виявились саме ті частотні складові, які адекватно відобразять частотний склад сигналу.

Існує континуум можливих частот складових в довільному сигналі, кожна з яких може бути присутньою в ньому.

Отже, ми маємо два варіанти:

1) в сигналі присутні лише такі гармонічні складові, які є в тому базисі, що ми будемо використовувати для аналізу. Це значить, що сигнал складається з коливань на гармонічних частотах, які кратні основній частоті ω_{0s} ;

2) в сигналі можуть бути присутні коливальні складові на частотах, які не кратні основній частоті. В цьому випадку в базисі відсутня гармоніка, що має таку ж частоту, як коливальна складова в сигналі.

Розглянемо сигнал $z[n]$, який складається із суми сигналу $x[n]$ (який містить тільки коливальні складові на частотах, які рівні частотам базисних гармонічних функцій) та ще одного коливання на довільній частоті ω^* з амплітудою A :

$$z[n] = x[n] + A e^{j2\pi\omega^* n} = \sum_{m=0}^{N-1} c_x[m] e^{2\pi j \frac{m}{N} n} + A e^{j2\pi\omega^* n}.$$

Знайдемо його спектр:

$$\begin{aligned} c_z[m] &= \frac{1}{N} \sum_{n=0}^{N-1} z[n] e^{-2\pi j \frac{m}{N} n} = \frac{1}{N} \sum_{n=0}^{N-1} \left(x[n] + A e^{j2\pi\omega^* n} \right) e^{-2\pi j \frac{m}{N} n} = \\ &= \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-2\pi j \frac{m}{N} n} + \frac{1}{N} A \sum_{n=0}^{N-1} e^{j2\pi\omega^* n} e^{-2\pi j \frac{m}{N} n} = \\ &= c_x[m] + \frac{A}{N} \sum_{n=0}^{N-1} e^{2\pi j \left(\omega^* - \frac{m}{N} \right) n}. \end{aligned}$$

Видно, що відмінності в спектрах $c_z[m]$ та $c_x[m]$ будуть залежати від того, чому буде рівний другий доданок. В першому випадку, коли сигнал $z[n]$ складається лише із коливань на тих же частотах, що і частоти базисних гармонічних коливань, матимемо, що частота відокремленого коливання також співпадає з одною із частот:

$\omega_0 = \frac{m}{N}$. Отже, спектр матиме вигляд:

$$c_z[m] = c_x[m] + \frac{A}{N} \sum_{n=0}^{N-1} e^{2\pi j \left(\frac{m}{N} - \frac{m}{N} \right) n} = c_x[m] + \frac{A}{N} \sum_{n=0}^{N-1} 1 = c_x[m] + A.$$

Ця рівність буде виконуватися для всіх. Тобто, якщо частота коливання, присутнього в сигналі, є такою ж, як і частота гармоніки в базисі, зміни відбудуться лише для одної відповідної спектральної компоненти: вона стане більшою.

В другому випадку, коли частота коливання не співпадає з частотою жодної базисної функції $\omega_0 \neq \frac{m}{N}$, матимемо, що другий доданок буде присутній у виразах для коефіцієнтів $c_z[m]$ для будь-якого m . Тобто, якщо в сигналі присутнє хоча б одне коливання на частоті, якої немає серед частот базисних функцій, всі коефіцієнти в спектрі зазнають змін. Отже, навіть коефіцієнти, що відповідають складовим сигналу на частотах базисних функцій ($c_x[m]$) поміняються.

Отже, гармонічні складові, частоти яких співпадають з частотами базисних функцій, можуть давати проекції *лише на один* з векторів базису і таким чином спектр сигналу буде адекватний. Складові інших частот, які не співпадають з частотами гармонічних функцій, будуть давати ненульові проекції *на всі вектори базису*, змінюючи їх вклад в розклад сигналу. В цьому випадку отриманий спектр не буде адекватним реальному складу сигналу. На практиці це буде призводити до зміни значень коефіцієнтів розкладу та до хибного тлумачення спектру. Енергія тих складових, які не присутні в базисі, але присутні в сигналі, буде «перетікати» на всі ті складові, які в базисі є.

Внаслідок того, що спектр сигналу невідомий заздалегідь, невідомо, яку частоту дискретизації треба обрати і ділянку сигналу якої тривалості треба виміряти, щоб в результаті спектрального аналізу отримати складові саме на тих частотах, які присутні в спектрі насправді. Описане явище називається «розтіканням спектру» (spectral leakage, spectral smearing).

Розтікання спектру буде суттєво спотворювати спектр в тому випадку, якщо в сигналі присутня потужна складова на частоті, яка не є кратною основній частоті ω_0 . Для цієї складової немає відповідної гармоніки в базисі, в якому ведеться розклад, тому вся енергія, яку несе ця складова сигналу, «розтечеться» по всім іншим гармонікам, які в базисі присутні.

Властивістю та фундаментальним обмеженням спектрального аналізу за Фур'є є те, що в будь-якому розрахованому спектральному представленні сигналу будуть присутніми тільки ті гармонічні складові, частоти яких визначаються *лише* тривалістю сигналу та частотою його дискретизації. Для довільного сигналу

неможливо сказати, чи є насправді ці отримані гармонічні складові з даними енергетичними характеристиками в сигналі, чи даний спектр отримано в результаті розтікання енергії складових на деяких інших частотах. Від розтікання спектру неможна позбутись.

Отримане в результаті прямого перетворення Фур'є спектральне представлення сигналу буде точним в сенсі збереження енергії, тобто енергія сигналу в часовій області буде дорівнювати сумі енергій всіх гармонічних складових. Але якщо для того самого сигналу буде змінено, наприклад, тривалість зареєстрованої ділянки, то в базисі будуть вже гармонічні складові інших частот, їх кількість буде також іншою. Але дане ще одне представлення буде також точним. Для одного і того самого сигналу, вимірюного з різними частотами дискретизації та різної тривалості, можна отримати нескінченну кількість спектральних представлень, кожне з яких буде вірним та представляти енергію сигналу. Дійсне значення спектральних складових сигналу отримати в загальному випадку неможливо, воно може бути отримано лише для тих сигналів, спектральний склад яких відомий заздалегідь.

Отже, перетворення Фур'є довільного сигналу дасть вірний результат лише в тому випадку, коли сигнал складається з гармонічних складових на частотах, кратних основній частоті. Якщо ми маємо в сигналі коливний компонент з деякою

проміжною частотою, яка лежить між аналітичними частотами $\omega[m] = m \frac{\omega_s}{N}$, то

енергія цієї складової проявиться деякою мірою на всіх N частотах. Ми будемо мати спотворення, яке називається розтіканням спектру.

Раніше при обговоренні спектрального аналізу дискретних скінченних сигналів вже було зазначено, що інтервал між частотами гармонік в спектрі залежить від кількості відліків цього сигналу: чим більше відліків – тим менша різниця між частотами і тим більш «багатий» спектр. Треба ще раз наголосити, що лише збільшення тривалості вимірювання сигналу допоможе отримати складові на частотах, які лежать достатньо близько одна до одної.

Між розтіканням спектру та ефектом Гіббса існує зв'язок. Зрозуміло, що обраний проміжок вимірювання сигналу повинен містити один період гармоніки основної частоти та цілу кількість періодів гармонік всіх частот, кратних основній. Припустимо, що в сигналі будуть присутні складові, які відмінні від тих, що є в даному базисі. Це призводить до того, що на проміжку вимірювання сигналу не укладається ціла кількість періодів цих складових сигналу, а отже, значення на початку і в кінці інтервалу не будуть рівними.

Ефект Гіббса та розтікання спектру, є наслідком невдалого вибору параметрів перетворення при якому маємо спотворення спектру реального сигналу. В загальному випадку позбавитися від цього ефекту неможливо.

3.15. Аналіз сигналів з віконними функціями

Для зменшення розтікання спектру (позбавлення від ефекту Гіббса) треба усунути нерівність значень сигналу на початку та в кінці проміжку, на якому він заданий (інтервалу часу вимірювання). На практиці в реальному часі це зробити неможливо, оскільки перетворення Фур'є дискретного сигналу дасть нам правильний результат тільки у випадку, коли в проміжок аналізу потрапляє ціла кількість періодів синусоїдальних компонент з частотою, яка кратна основній частоті. Оскільки спектр заздалегідь невідомий, то і обрати оптимальний проміжок для розрахунку спектру неможливо.

Можна застосовувати інший підхід зменшення присутності небажаних складових в спектрі сигналу – використання спектрального аналізу з віконними функціями (зважування, перетворення Фур'є з вікном, віконної обробки, windowing).

Інтуїтивним шляхом позбавлення від ефекту Гіббса є забезпечення рівності значень сигналу на початку і в кінці проміжку вимірювання. Для цього перед тим, як розраховувати перетворення Фур'є сигнал перемножається на деяку вагову (віконну) функцію $w(t)$, яка повинна спадати біля кінців того сегменту, на якому ми розглядаємо сигнал (рис. 3.19, б). в середині сегменту вона повинна мати постійні значення.

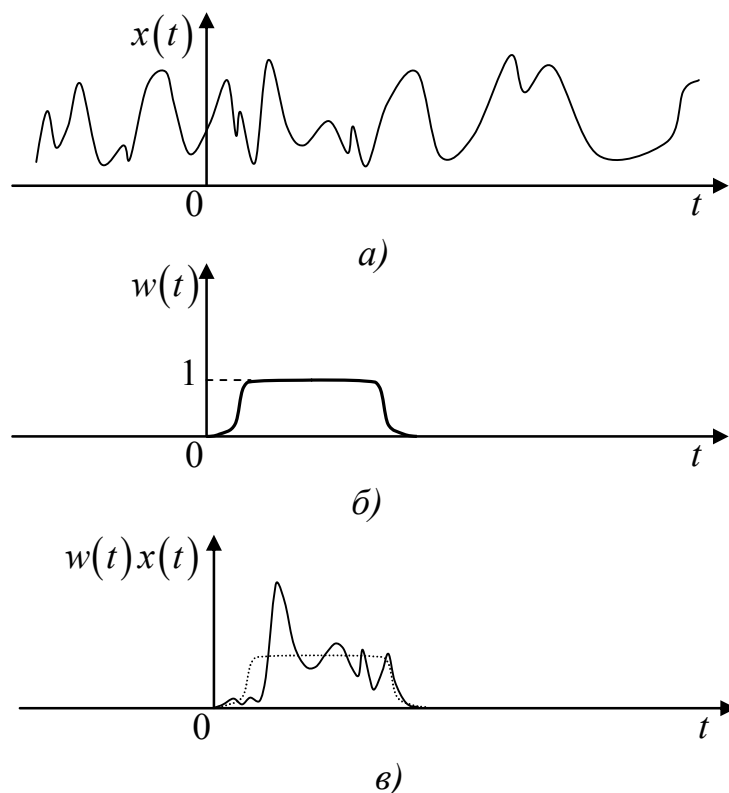


Рисунок 3.19 – а) досліджуваний сигнал, б) графік віконної функції, в) результат перемноження сигналу та вікна

В результаті перемноження вікна на сигнал отримаємо ділянку сигналу, яка розташована на тому інтервалі часу, який потрапляє у вікно $w(t)$. Видно (рис. 3.19, г), що ця ділянка на краях є дещо спотвореною, але в середині інтервалу співпадає з початковим сигналом. До того ж, значення сигналу в кінці і на початку інтервалу рівні одне одному (нульові), що і вимагається для позбавлення від ефекту Гіббса. Вираз для прямого перетворення Фур'є в цьому випадку матиме вигляд:

$$F_w(\omega) = \int_{-\infty}^{\infty} x(t)w(t)e^{-j\omega t} dt, \quad (3.37)$$

$$c_w[m] = \sum_{n=0}^{N-1} x[n]w[n]e^{-2\pi j \frac{m}{N}n}. \quad (3.38)$$

Авжеж, не можна сказати, що використання віконних функцій приведе до того, що отриманий за (3.37) або (3.38) спектр буде таким самим, як спектр початкового сигналу. В вирази для отримання спектру замість ділянки сигналу, яку ми хочемо проаналізувати, входить цей сигнал, домножений на віконну функцію, а отже спотворений внаслідок перемноження. Зміна форми сигналу в часовій області не може не привести до зміни його спектральних характеристик. З властивостей перетворення Фур'є відомо, що множення сигналів в часовій області призводить до того, що спектр результуючого сигналу буде дорівнювати згортці спектрів початкових сигналів:

$$F_w(\omega) = X(\omega) * W(\omega),$$

де $X(\omega)$ —спектр сигналу;

$W(\omega)$ —спектр віконної функції.

Отже, спектр сигналу, який ми отримаємо після операції зважування, буде зміненим по формі.

З міркувань мінімізації спотворення внаслідок згортання спектрів сигналу та віконної функції при зважуванні, впливає вимога до віконних функцій. А саме, їх спектри повинні бути максимально вузькими та сконцентрованими в околі точки $\omega = 0$.

Питання вибору оптимальної віконної функції не вирішено зараз остаточно. Існує багато вікон, деякі з них використовуються частіше, деякі є достатньо мало розповсюдженими. Розглянемо деякі основні дискретні віконні функції (рис. 3.20).

Прямокутне вікно (вікно Діріхле):

$$w[n] = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \text{в інших випадках} \end{cases}$$

Основне вікно, яке використовують для аналізу, і те вікно, яке ми неявно використовували досі, коли говорили про Фур'є-аналіз дискретних скінченних сигналів. Можна вважати, що сигнали, які зареєстровані на скінченному проміжку

часу, отримані в результаті добутку прямокутної віконної функції на нескінченний сигнал.

Трикутне вікно (вікно Бартлета):

$$w[n] = \begin{cases} \frac{n}{N/2}, & 0 \leq n \leq \frac{N}{2} \\ w[N-n], & \frac{N}{2} < n \leq N-1 \end{cases}$$

Родина вікон $\cos^\alpha(X)$

Це вікна, що отримуються за загальним виразом

$$w[n] = \cos^\alpha \left[\pi \frac{n}{N} \right], \quad -\frac{N}{2} \leq n \leq \frac{N}{2}.$$

Як правило, показник степеню є цілим числом, найчастіше використовують значення α від 1 до 4. Перевагою використання цих вікон є порівняна простота розрахунку значень віконної функції для різних тривалостей. Найбільш розповсюдженим є випадок $\alpha = 2$ (**вікно Хенна**):

$$w[n] = \begin{cases} \cos^2 \left(\pi \frac{n}{N} \right) = \frac{1}{2} \left(1 + \cos \left(2\pi \frac{n}{N} \right) \right), & -\frac{N}{2} \leq n \leq \frac{N}{2} \\ 0, & \text{в інших випадках} \end{cases}$$

Вікно Хеммінга:

$$w[n] = \begin{cases} \alpha + (1-\alpha) \cos \left(2\pi \frac{n}{N} \right), & -\frac{N}{2} \leq n \leq \frac{N}{2} \\ 0, & \text{в інших випадках} \end{cases}$$

При значенні $\alpha = \frac{25}{46} \approx 0.54$ спектр даного вікна має оптимальні характеристики з точки зору співвідношення основного та бічних пелюстків.

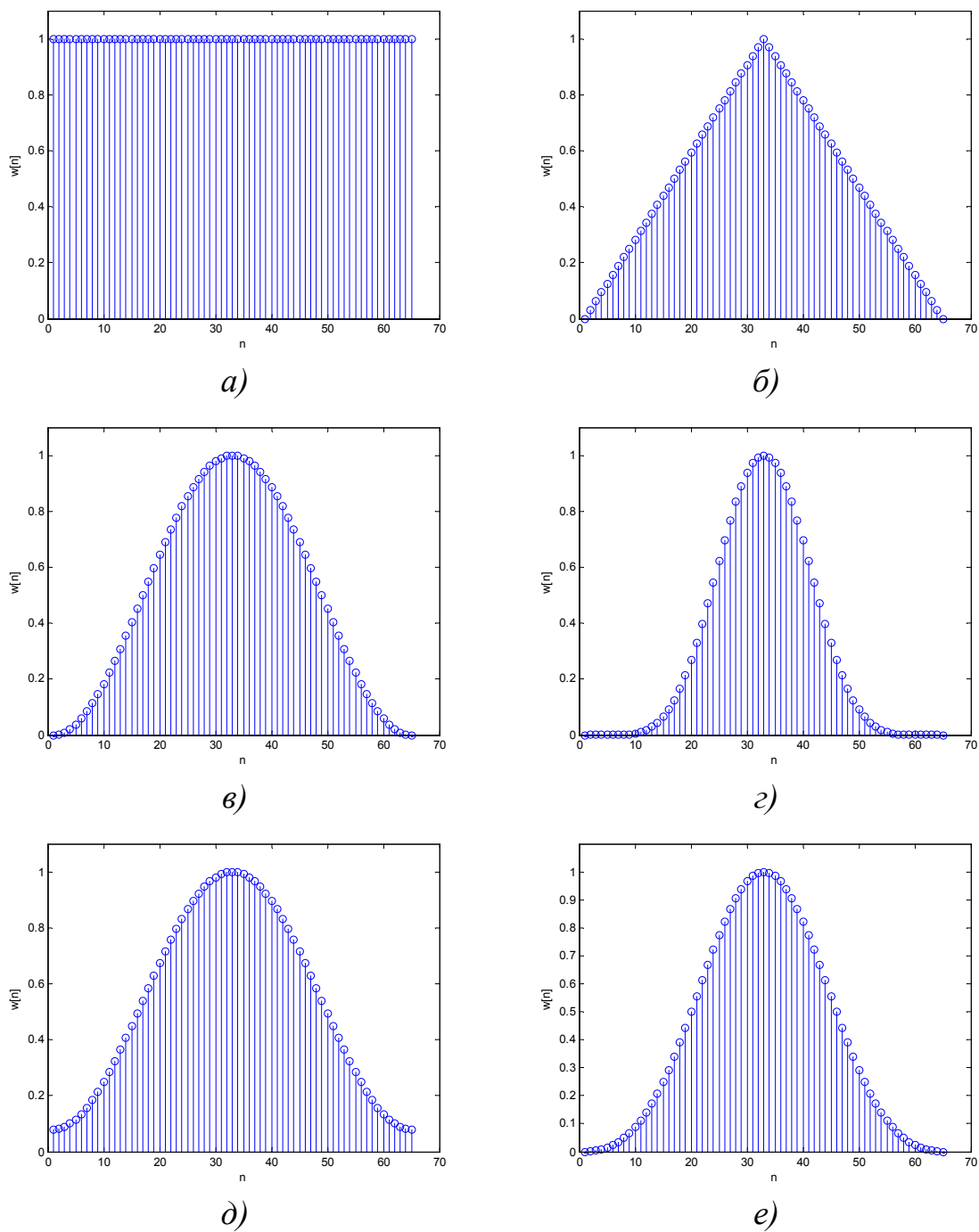


Рисунок 3.20 – Віконні функції а) прямокутне вікно, б) трикутне вікно, в) вікно Хенна, г) вікно $\cos^6(X)$, д) вікно Хеммінга, е) вікно Блекмена,

Вікно Блекмена:

$$w[n] = \begin{cases} 0.42 + 0.5 \cos\left(2\pi \frac{n}{N}\right) + 0.08 \cos\left(4\pi \frac{n}{N}\right), & -\frac{N}{2} \leq n \leq \frac{N}{2} \\ 0, & \text{в інших випадках} \end{cases}$$

Вікно Ріца:

$$w[n] = 1 - \left| \frac{n}{N/2} \right|^2, \quad 0 \leq |n| \leq \frac{N}{2}.$$

Вікно Тьюкі:

$$w[n] = \begin{cases} 1, & 0 \leq |n| \leq \alpha \frac{N}{2} \\ 0.5 \cdot \left(1 + \cos \left(\pi \frac{n - \alpha \frac{N}{2}}{2(1 - \alpha) \frac{N}{2}} \right) \right), & \alpha \frac{N}{2} \leq |n| \leq \frac{N}{2} \end{cases}$$

Це вікно змінює свою форму від прямокутного вікна до вікна Хенна при зміні величини параметра α від нуля до одиниці.

Серед інших вікон використовують вікна Пуассона, Гауса, Бомена, Ла Валя, Чебишова та інші. Вибір вікна залежить від параметрів спектру вікна та від того, наскільки спектр аналізованого сигналу буде спотворений при процедурі зважування.

3.16. Підсумки до розділу 3 та рекомендована література

В цьому розділі було розглянуто теорію класичного і до цього часу найбільш розповсюдженого методу аналізу сигналів – спектрального аналізу за Фур'є. Він полягає у розкладі сигналу на гармонічні складові, неперервні або дискретні залежно від того, які сигнали аналізуються. Спектр сигналу дає досліднику уяву про те, коливання на яких частотах складають сигнал.

Детально досліджено, що відбувається зі спектром сигналу при його дискретизації. Визначено, що спектр дискретного сигналу стає періодичним та складається з копій спектру початкового неперервного сигналу. Оскільки періодом повторення копій спектру є частота дискретизації, то якщо вона обрана надто низькою, то можливе явище накладання спектрів. В цьому випадку отриманий дискретний сигнал більше

не буде мати всі властивості початкового неперервного. Правило для вибору частоти дискретизації дає теорема Котельникова-Шеннона.

Не всі сигнали мають спектр, який сконцентрований біля нульової частоти. Для їх дискретизації використовують, зокрема, змогову дискретизацію. Для дискретизації вузькосмугових високочастотних сигналів частота дискретизації обирається за правилом, яке забезпечує ненакладання копій спектрів, а не самого спектру. Це дає змогу ослабити вимоги до аналого-цифрового перетворювача.

Розглянуто також ще два тісно пов'язаних явища, що відбуваються при дискретизації, а саме, розтікання спектру та явище Гіббса. Для уникнення останнього можна використовувати віконні функції. При віконній обробці спектральному аналізу підлягає не сам сигнал, а його добуток з ваговою функцією. Для більш глибокого ознайомлення зі спектральним аналізом за Фур'є можна скористатися [2, 21-23]. Особливості спектрального аналізу саме дискретних сигналів висвітлені в [2-4, 15]. Питання віконної обробки висвітлені в [7], а про властивості різних вікон можна прочитати в ґрунтовній статті [24].

3.17. Питання для самоперевірки

1. За якими виразами отримують спектральну густину та спектральну густину потужності для: неперервних періодичних нескінченних сигналів, неперервних неперіодичних сигналів, дискретних нескінченних сигналів, дискретних скінченних сигналів?
2. Які властивості перетворення Фур'є для неперервних та дискретних сигналів?
3. Що відбувається зі спектром сигналу при дискретизації? Яким математичним виразом описується зв'язок між спектрами початкового неперервного сигналу та отриманого дискретного сигналу?
4. В чому полягає явище накладання спектрів при дискретизації, як його уникнути? Сформулювати теорему Котельникова та її наслідки.
5. Які особливості субсмугової дискретизації? Як обирається частота дискретизації у випадку вузькосмугових сигналів?
6. В чому полягає явище розтікання спектру?
7. В чому полягає ефект Гіббса, як можна його уникнути?
8. Як проводять віконну обробку сигналів, яку користь вона може принести? Які віконні функції бувають?

4. СПЕКТРАЛЬНО-ЧАСОВИЙ ТА ІНШІ ВИДИ АНАЛІЗУ СИГНАЛІВ

4.1. Спектрально-часовий аналіз за Фур'є

Нехай маємо достатньо тривалий сигнал $x(t)$ (рис. 4.1а), і поставимо задачу проаналізувати його спектральний склад. Для цього треба взяти деяку ділянку цього сигналу, залучити віконну функцію та скористатися прямим перетворенням Фур'є. Видно, що в сигналі $x(t)$ одночасно присутні складові частини, які змінюються достатньо швидко (і будуть описані в спектральному розкладі гармоніками високих частот), а також повільні складові (яким будуть відповідати низькочастотні гармонічні складові в розкладі). Ці складові в різних частинах обраного нами вікна можуть мати різні амплітуди. Отже, якщо для розрахунку спектру буде взято весь сигнал одночасно, то його спектр буде достатньо широким, в ньому будуть складові і на верхніх, і на нижніх частотах (рис. 4.1б).

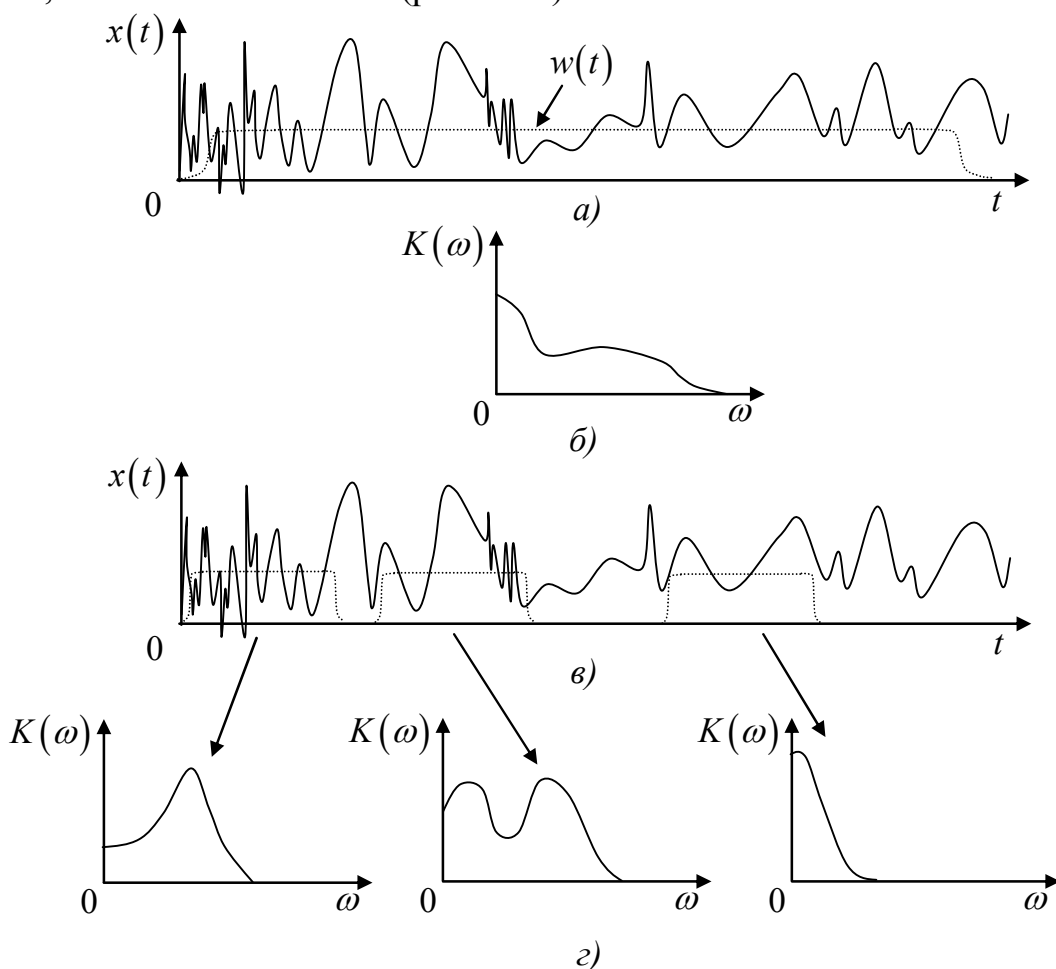


Рисунок 4.1 – а) сигнал, спектр якого змінюється протягом спостереження, б) спектр, розрахований по всьому сигналу, в) можливі варіанти потрапляння у вікна ділянок сигналу з різним спектром, г) спектри ділянок сигналу

Можна побачити, що на тій ділянці, яка потрапила у широке вікно, характер сигналу змінюється достатньо суттєво. На початку ділянки в сигналі переважають швидкі складові, які мають достатньо значну величину. Якби у більш вузьке вікно потрапила лише ця частина сигналу (рис. 4.1в), то спектр був би суттєво відмінним від спектру всього сигналу (рис. 4.1г), і сконцентрованим на верхніх частотах. Далі властивості сигналу ще змінюються (рис. 4.1в), і в ньому з'являються високоамплітудні повільні складові, але через деякий час присутній сплеск високоамплітудної високочастотної активності. Якби з допомогою віконної функції було виділено цю ділянку, спектр відобразив би наявність потужних складових на нижніх та верхніх частотах. Згодом в сигналі пропадають високочастотні коливання, і сигнал стає повільним. Спектр цієї частини мав би містити лише складові на нижніх частотах. Видно, що залежно від розташування віконної функції в спектрі будуть присутніми ті чи інші компоненти, оскільки в сигналі з'являються чи пропадають ділянки швидких або повільних змін.

Отже, в загальному випадку спектральний склад сигналу може змінюватися навіть протягом нетривалого спостереження. Така зміна спектру пов'язана з тим, що змінюються характеристики сигналу, а це значить, що змінюється характер процесів, які відбуваються в об'єкті спостереження, від якого цей сигнал зареєстровано. Ці зміни становлять значний інтерес для дослідника, бо вони можуть бути ознаками, які несуть інформацію про об'єкт. Важливо знати, яким є спектр сигналу протягом певного періоду часу, і в який саме момент характеристики сигналу змінилися. Спостереження за динамікою об'єкта на основі аналізу змін в часі спектральних характеристик сигналів – одна з найважливіших задач, на вирішення якої направлені методи теорії сигналів. Такого роду задачі з'являються в радіолокації, системах зв'язку, моніторингу, сегментації біомедичних сигналів та ін. Дивлячись на спектр всього сигналу (рис. 4.1б), не можна сказати, в який саме момент часу в сигналі з'явилася чи зникла складова тої чи іншої частоти. При класичному спектральному аналізі за Фур'є вважається, що кожна гармонічна складова присутня в сигналі в усі моменти часу, коли існує сигнал, тобто, що спектральний склад не змінюється з часом. Отже, спектральне представлення за Фур'є втрачає часову інформацію про спектральні складові.

Для аналізу сигналів зі змінним спектром потрібне інше представлення сигналів – двовимірне частотно-часове представлення. Необхідно мати можливість знати одночасно і спектральний склад, і тривалості присутності в сигналі складових на різних частотах.

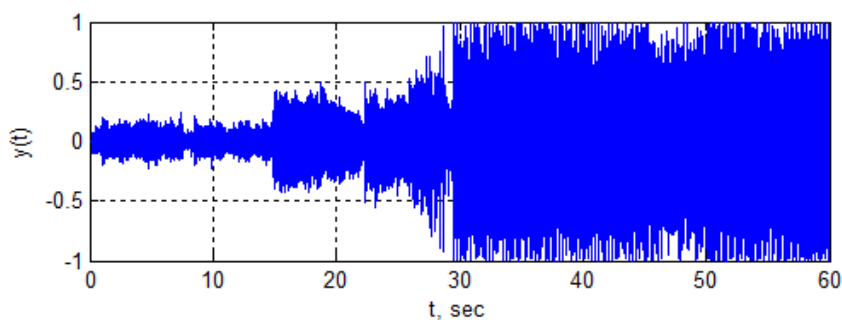
Прикладом такого представлення є музикальні ноти: по горизонтальній вісі відкладається час, а по вертикальній вказується, тони яких частот повинні бути присутні у сигналі в даний момент часу, та скільки вони мають тривати. На рис. 4.2а наведені ноти для пісні, а на рис. 4.2б – графік перших шестидесяти секунд одного з каналів звукозапису цього твору («Любит. Не Любит. Я Руки Ломаю», вокально-інструментальний ансамбль «Date Rape»).

Інтуїтивно зрозуміло, що для отримання інформації про зміну спектру з часом треба переміщати віконну функцію по сигналу, і розраховувати спектр для тих ділянок, які потрапили у вікно. Набір таких спектрів буде розглядатися як єдине ціле для описання залежності спектрального складу від часу.

Dropped D Tuning
 ① = E ④ = D
 ② = B ⑤ = A
 ③ = G ⑥ = D

Moderate $\text{♩} = 130$

а)



б)

Рисунок 4.2 – Частотно-часове представлення звукового сигналу у вигляді нот (а), графік запису звукового сигналу (б)

Математичний вираз для спектрально-часового розкладу сигналу називають короткочасовим перетворенням Фур'є (Short-Time Fourier Transform). Розрахунок спектрально-часового представлення сигналів проводиться таким чином. Спочатку віконну функцію $w(t)$ розташовують в околі точки $t=0$. Після цього розраховується спектр за Фур'є добутку $x(t)w(t)$. Потім віконну функцію зміщують на певний крок τ_1 , формують функцію $w(t-\tau_1)$ і розраховують спектр добутку $x(t)w(t-\tau_1)$. Далі процедура повторюється для всіх можливих зміщень віконної функції. Вирази спектрально-часового розкладу неперервних та дискретних сигналів мають вигляд:

- в неперервному випадку:

$$F_w(\omega, \tau) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j\omega t} dt, \quad (4.1)$$

- дискретному випадку:

$$c[m, k] = \sum_{n=0}^{N-1} x[n]w[n-k]e^{-2\pi j \frac{m}{N} n}, \quad (4.2)$$

де $w(t), w[n]$ – функція вікна;

τ, k – параметр зміщення віконної функції в часі.

Представлення розподілу потужності спектральних складових, яке відображає зміну спектру в часі, називається *спектрограмою*:

$$S_w(\omega, \tau) = \left| \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j\omega t} dt \right|^2, \quad (4.3)$$

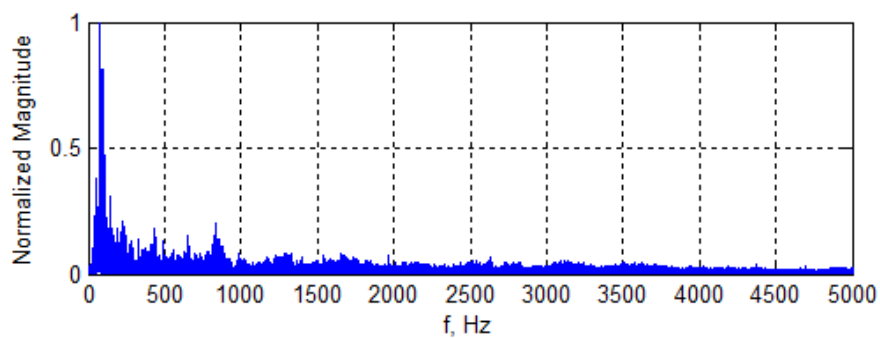
$$C_w[m, k] = \left| \sum_{n=0}^{N-1} x[n]w[n-k]e^{-2\pi j \frac{m}{N} n} \right|^2. \quad (4.4)$$

Для прикладу можна навести результати спектрального та спектрально-часового аналізу одного і того самого сигналу за Фур'є (рис. 4.3). На рис. 4.3а наведено графік амплітудного спектру звукового сигналу з рис. 4.2б; можна помітити, що в спектрі присутні коливання на різних частотах, дещо більше енергії зосереджено в проміжку частот до 500 Гц. Але можна також помітити, що в проміжку від 1 кГц до 4 кГц амплітуди спектральних складових приблизно однакові, отже, енергія розподілена рівномірно. Коли саме присутні коливання на різних частотах, з цього спектру побачити неможливо.

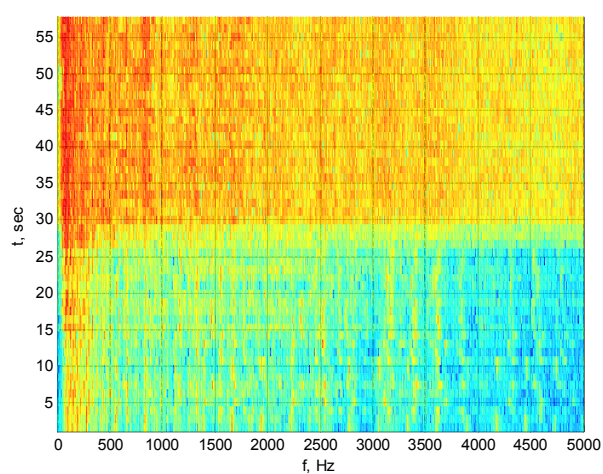
Якщо подивитись на спектрограму сигналу (рис. 4.3б), то можна отримати більше інформації. На спектрограмі значення амплітуд гармонічних складових $C_w[m, k]$ визначаються кольором: тут великим значенням відповідає червоний, меншим – жовтий, зовсім малим – синій. Видно, в які проміжки часу і які коливання в спектрі присутні. Зокрема, можна помітити, що на початку сигналу спектральні складові присутні в основному на частотах від 100 до 300 Гц (в цьому проміжку багато областей жовтого і червоного кольору). Від 500 Гц до 3.5 кГц в спектрограмі

присутні складові і великої, і малої величини, однакові для всього проміжку; спектральні складові на частотах більше 4 кГц майже відсутні (амплітуди зображені здебільшого синім кольором, що відповідає малим значенням). В пісні цей час відповідає гітарному вступу.

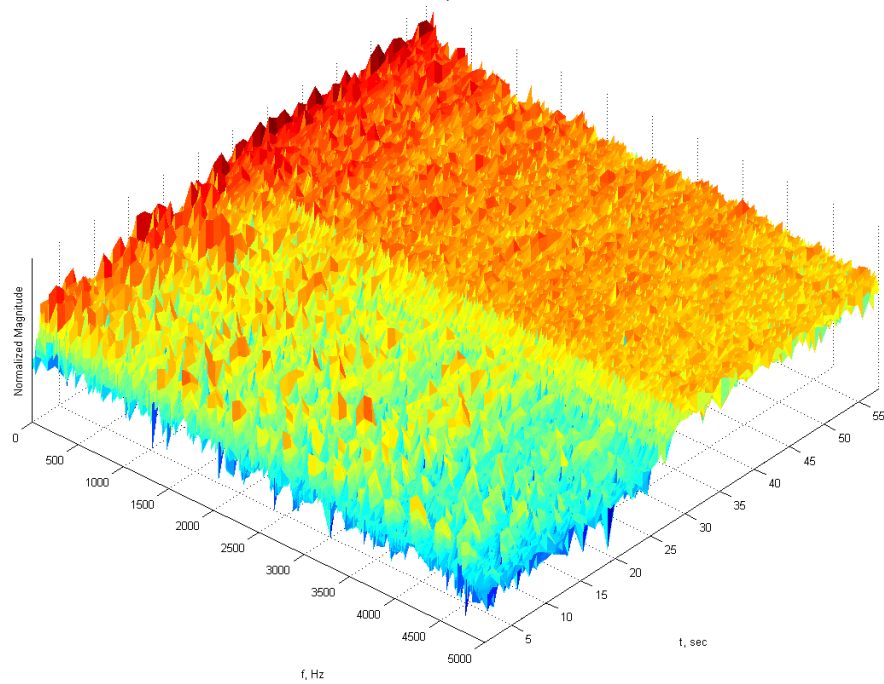
Починаючи з 30-ї секунди спектральний склад сигналу змінюється: видно, що значно збільшується амплітуда складових в проміжку до 1 кГц, і дещо меншою, все одно порівняно великою стають амплітуди складових від 1 до 5 кГц. Спектр сигналу стає високоамплітудним і широким. В пісні в цей час вступають кілька гітар, голос вокаліста та ударні. На рис. 4.3в спектрограма зображена у вигляді графіку поверхні, на якій можна спостерігати зміну значень амплітуд спектральних складових.



а)



б)



в)

Рисунок 4.3 – Результати аналізу звукового сигналу за Фур'є: спектр (а), спектрограма (б), тривимірне зображення спектрограми (в)

Отже, для аналізу реальних сигналів, як правило, необхідно застосовувати частотно-часові перетворення, які дають змогу спостерігати за зміною спектрального складу сигналу у часі.

Питаннями, які треба вирішити до практичної реалізації частотно-часового аналізу є вибір типу вікна, тривалості вікна, перекриття вікон та кроку зміщення. Залежно від того, якими будуть ці параметри, будуть отримані ті чи інші результати навіть для одного і того самого сигналу.

Першим прикладом спектрально-часового аналізу було *перетворення Габора* (запропоновано інженером Деннісом Габором у 1946 році). Воно відрізняється тим, що в якості віконної функції використовується функція Гаусса. Вираз (4.1) набуває вигляду:

$$\begin{aligned} G(\omega, \tau) &= \int_{-\infty}^{\infty} x(t) h_a(t - \tau) e^{-j\omega t} dt = \\ &= \int_{-\infty}^{\infty} x(t) e^{-\pi(t-\tau)^2} e^{-j\omega t} dt. \end{aligned} \quad (4.5)$$

Перевагами перетворення Габора є те, що використання таких вікон дозволяє оптимально локалізувати всі нерегулярності та особливості сигналу в часовій та частотній області. Після перетворення Габора дослідники почали використовувати інші віконні функції для частотно-часового аналізу.

Введення в практику аналізу сигналів короткочасового перетворення Фур'є стало суттєвим кроком до більш адекватного дослідження сигналів, спектр яких змінюється в часі. Задаючи віконну функцію певної ширини та зміщуючи її по часовій вісі, можна отримати достатньо докладний опис особливостей спектру сигналу на різних ділянках. Але обмеженням, яке має такий спектрально-часовий аналіз, є те, що ширина вікна, яке використовується при аналізі, є постійною для всього сигналу. Отже, незалежно від того, які характеристики має сигнал, і на яких ділянках його спектр має постійні параметри, вікно буде одним і тим самим. В реальних сигналах може бути ситуація, коли складові частини з різними частотами можуть мати різну тривалість. І не завжди тривалість ділянки сигналу, на якій спектр буде постійним, співпадає з тривалістю віконної функції. Тим більше, не завжди це відбувається коли аналіз сигналу проводиться в реальному часі.

Мати вікно постійної ширини не дуже зручно, оскільки в цій ситуації також не можна бути впевненими, що при зміщенні віконної функції у вікно потрапить саме ділянка сигналу, на якій спектральний склад постійний. Наприклад, на рис. 4.1, г, видно, що у другому вікні на початку ділянки сигнал містить повільне колювання, а в кінці цього вікна в сигналі з'являються потужні високочастотні колювання. Тобто спектральний склад частини сигналу на початку вікна та в кінці вікна явно відрізняється. Але вони опинилися в одному вікні, тому по характеристикам спектру не можна буде розрізнити, в які саме моменти які частини сигналу присутні на даній ділянці.

Для того, щоб уникнути такої ситуації, необхідне подальше удосконалення спектрально-часового аналізу. Для того, щоб кожна ділянка сигналу була аналізована у вікні з оптимальною шириною, треба мати можливість зміни тривалості (масштабування в часі) віконної функції. Тобто, в самому виразі для розкладу сигналу тривалість вікна повинна бути присутньою як параметр. В рамках класичного спектрального аналізу це зробити не можна, тож треба залучати розклади сигналу на складові, які відмінні від гармонічних.

4.2. Перетворення Хаара

Перший в світі різновид розкладу сигналу на складові різної тривалості був запропонований угорським математиком Альфредом Хааром у 1910 році. Він запропонував перетворення сигналу, в якому розклад ведеться не на синусоїдальні складові, а на сукупність біполярних прямокутних імпульсів, які зміщуються по вісі часу та мають різну ширину.

Будемо розглядати Гілбертів простір неперервних сигналів зі скінченною енергією, заданих на одиничному інтервалі: $x(t) \in L^2(0,1)$. Обмеження на область визначення введена лише для зручності математичного описання, оскільки одиничний інтервал можна масштабувати (розтягнути або стиснути) в довільне число разів, а також змістити по числовій вісі. При цьому можна аналізувати сигнали, які задані будь-де на числовій вісі часу, ввівши перетворення лише на інтервалі від 0 до 1.

Введемо до розгляду «материнську функцію» Хаара (рис. 4.4):

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1 \\ 0, & \text{в інших випадках} \end{cases}$$

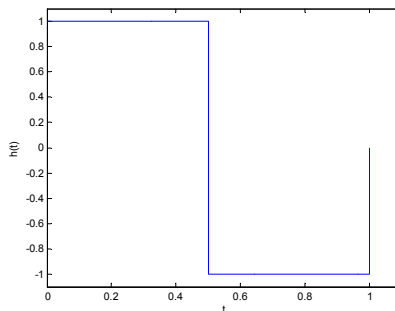


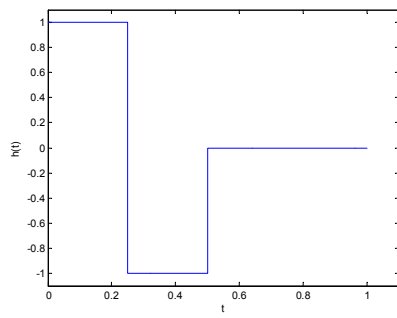
Рисунок 4.4 – Материнська функція Хаара

На основі цієї материнської функції сформуємо набір функцій за таким правилом:

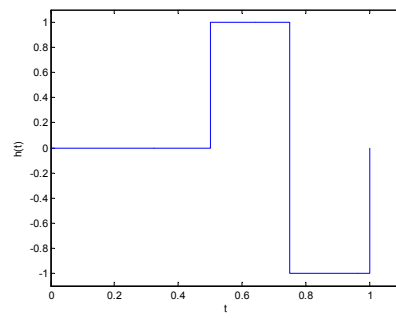
$$\psi_{n,k}(t) = \psi(2^n t - k), \quad n = 0, 1, 2, \dots, \quad 0 \leq k < 2^n, \quad (4.6)$$

де коефіцієнт n відповідає масштабу функції, а k – зміщенню функції в межах інтервалу $(0,1)$. Видно, що всі функції в наборі будуть масштабованими версіями материнської функції: чим більшим буде коефіцієнт 2^n , тим більш стиснутою буде функція. Також стиснуті функції зі створеного нами набору можуть бути зміщеними по вісі часу відносно початку координат відповідно до параметра k . Приклади стиснутих та зміщених функцій наведені на рис. 4.5.

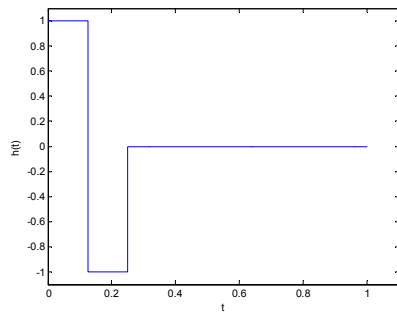
Треба звернути увагу на те, що для кожного масштабного коефіцієнта n якщо розглянути всі можливі зміщення масштабованої функції Хаара, то вони покриють собою всю числову вісь від 0 до 1. Іншими словами, перебираючи всі k , можна розташувати функцію у будь-якій частині інтервалу. Наприклад, таке буде для функцій на рис. 4.5, в-е: всі чотири функції мають ненульові значення на різних ділянках проміжку від 0 до 1.



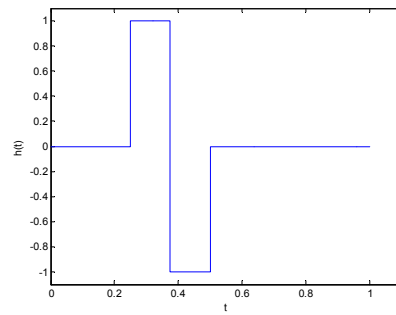
а)



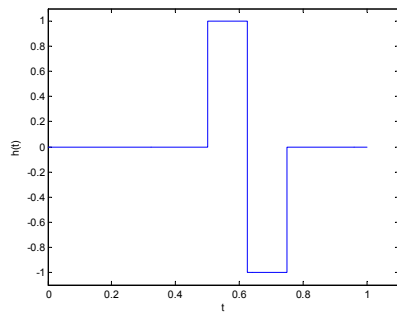
б)



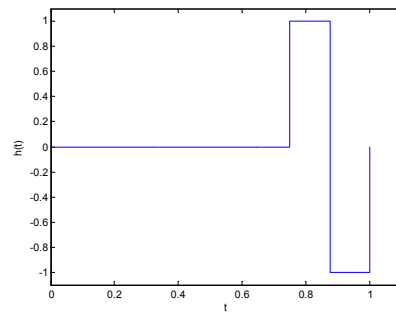
в)



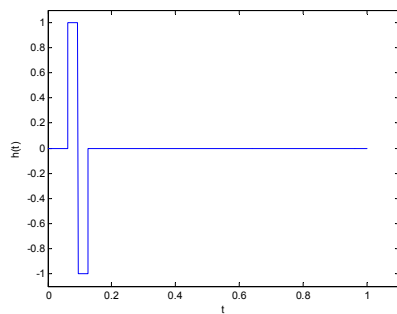
г)



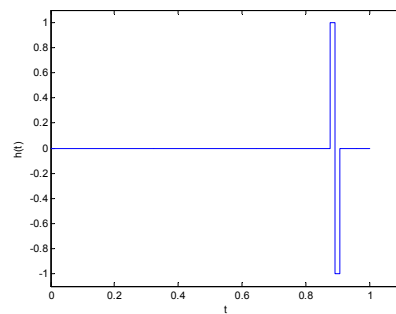
д)



е)



ж)



з)

Рисунок 4.5 – Базисні функції Хаара: а) $\psi_{1,0}(t) = \psi(2t)$, б) $\psi_{1,1}(t) = \psi(2t-1)$, в) $\psi_{2,0}(t) = \psi(4t)$, г) $\psi_{2,1}(t) = \psi(4t-1)$, д) $\psi_{2,2}(t) = \psi(4t-2)$, е) $\psi_{2,3}(t) = \psi(4t-3)$, ж) $\psi_{4,1}(t) = \psi(16t-1)$, з) $\psi_{5,28}(t) = \psi(32t-28)$

Покриття локалізованими в часі базисними функціями всієї області визначення сигналу – ключова вимога для того, щоб аналіз сигналу, характер якого змінюється в часі, був адекватний. Коли в базисі присутні функції, які можуть бути розташовані будь-де на часовій вісі, то це значить, що у розкладі будуть представлені всі локальні особливості поведінки сигналу, де б у часі вони не знаходились.

Можна довести, що функції (4.6) на одиничному інтервалі складають повну ортогональну систему функцій (базис), який називається базисом Хаара. Отже, кожний неперервний сигнал, який заданий на одиничному інтервалі, може бути представлений у вигляді розкладу за таким базисом. Коефіцієнти розкладу (проекції сигналу на базисні функції) отримують так само, як і для будь-якого ортогонального базису в гільбертовому просторі, а саме як скалярні добутки, а загальний вираз розкладу сигналу матиме вигляд:

$$x(t) = \sum_{n,k} \frac{\langle x(t), \psi_{n,k}(t) \rangle}{\|\psi_{n,k}(t)\|^2} \psi_{n,k}(t).$$

Для отримання пари прямого та оберненого перетворень Хаара розрахуємо спочатку норму базисної функції для довільних параметрів n та k :

$$\begin{aligned} \|\psi_{n,k}(t)\| &= \sqrt{\int_0^1 \psi_{n,k}(t) \cdot \psi_{n,k}^*(t) dt} = \sqrt{\int_0^1 \psi^2(2^n t - k) dt} = \\ &= \sqrt{\frac{1}{2^n} \int_0^1 \psi^2(2^n t - k) d(2^n t - k)} = \sqrt{\frac{1}{2^n} \left[\int_0^{0.5} 1^2 dt + \int_{0.5}^1 (-1)^2 dt \right]} = \frac{1}{\sqrt{2^n}} \end{aligned}$$

Пряме перетворення Хаара буде задаватися виразом:

$$H(n, k) = \langle x(t), \psi_{n,k}(t) \rangle = \frac{1}{2^n} \int_0^1 x(t) \cdot \psi_{n,k}(t) dt, \quad (4.7)$$

Обернене перетворення Хаара буде мати вигляд:

$$x(t) = \sum_{n=0}^{+\infty} \sum_{k=0}^{2^n} H(n, k) \psi_{n,k}(t). \quad (4.8)$$

Воно відображає розклад сигналу: представлення $x(t)$ у вигляді суми складових, кожна з яких є масштабованою та зміщеною копією материнської функції Хаара.

Отже, отримано розклад сигналу на масштабовані біполярні імпульси (функції Хаара). В цьому розкладі збережена локальна інформація про властивості сигналу, оскільки базисні функції визначені не на всій числовій вісі одночасно, а є локалізованими в часі. Недоліком розкладу сигналу на функції Хаара є те, що ці функції є розривними (містять різкі зміни значень). А отже, за допомогою базису Хаара буде складно представити гладкі сигнали без розривів.

Перетворення Хаара широко використовується при обробці та аналізі зображень.

Варто зазначити, що функція Хаара є найпростішим вейвлетом, а перетворення Хаара є історично найпершим вейвлет-перетворенням.

4.3. Вейвлет-перетворення

Як було показано у попередніх розділах, один з корисних способів аналізу сигналів, характеристики яких (в тому числі спектральні) змінюються в часі, є частотно-часовий розклад за допомогою віконних функцій. При цьому виникають деякі проблеми:

- заздалегідь невідомо, яка тривалість ділянки сигналу матиме постійний спектр, тобто ми не знаємо, яка тривалість віконної функції нам потрібна;
- невідомо, наскільки швидко змінюється спектральний склад сигналу, тобто ми не знаємо, з яким кроком треба переміщати вікно аналізу;
- якщо віконна функція буде дуже вузькою, у вікно потрапляє мало сигналу. Відомо, що інтервал між гармоніками в спектрі обернено пропорційний довжині ділянки сигналу, яка аналізується, отже, якщо ділянка маленька, то в спектрі ми зможемо розрізнити лише невелику кількість частотних компонентів. Тому беручи короткі вікна ми тим самим погіршуємо спектральну роздільну здатність. Якщо вікно розширити – туди будуть потрапляти тривалі ділянки сигналу, і тому ми погіршуємо здатність такого аналізу по визначенню локальних характеристик сигналу. Тобто при збільшенні тривалості вікна погіршується роздільна здатність в часі.

З метою отримання адекватного спектру до тривалості віконної функції висуваються одночасно дві протилежні вимоги: ділянка сигналу, яка вирізається вікном, повинна бути настільки малою, щоб його спектральні характеристики можна було вважати постійними на ній. Але водночас вікно повинно бути настільки широким, щоб в спектрі сигналу ми були здатні розрізнити багато гармонік.

Можна подивитись на цю проблему по-іншому: відомо, що чим нижча частота, тим триваліший період коливання. Тому щоб у вікно потрапило багато відліків «складових» низької частоти, треба мати тривалу ділянку сигналу (тривале вікно). Швидкі компоненти сигналу описуються гармоніками високої частоти, які мають короткий період. Тому щоб їх зареєструвати, достатньо коротко часового вікна. Видно, що для визначення спектральних складових різних частот треба мати вікна різної тривалості. Отже, потрібен компроміс, і оптимальним рішенням може бути аналіз сигналу одночасно з використанням коротких ділянок та тривалих ділянок сигналу. В цьому випадку можна розкласти сигнал зі змінними локальними властивостями на гармонічні складові різної тривалості та різних частот. Класичний короткочасовий аналіз за Фур'є не дозволяє цього зробити.

Для розкладання сигналів на складові частини різної тривалості і з різним спектральним складом був розроблений новий інструментарій аналізу сигналів – *вейвлет-перетворення*.

Вейвлет-перетворення (від слова *wavelet* – маленька хвиля), введено в теорію сигналів у 1984 році в роботі геофізика Жана Морле та математика Александра Гросмана “Розклад функцій Харді у інтегровані з квадратом вейвлети сталої форми”, в якій вони започаткували напрямок розкладу сигналів за допомогою т.зв. фреймів. З того часу “вейвлетний” підхід успішно розповсюдився у численні напрямки аналізу сигналів – знешумлення, субсмугова фільтрація, кодування в цифровому зв’язку, маркування та стиснення багатовимірних сигналів, детектування сингулярностей, розпізнавання образів, обробка електроенцефалограм та аналіз варіабельності ритму серця, аналіз та обробка біомедичних зображень. Відомі вчені-математики та інженери працюють над розвитком та вдосконаленням теорії та практичного застосування вейвлетів і зараз.

Загальноприйняте класичне визначення вейвлет-перетворення дала відома математик Інгрід Добеші у своїй роботі 1992 року “Десять лекцій про вейвлети”: “Вейвлет-перетворення є інструментом, що розбиває дані, або функції, або оператори на складові з різним частотним складом, і потім досліджує кожну із роздільною здатністю, яка відповідає її розміру”.

По суті, вейвлет-аналіз – це розклад сигналів на складові різної форми та тривалості. Розглянемо принципи використання вейвлет-перетворень для аналізу сигналів. Як і будь-який розклад, вейвлет-розклад аналізованого сигналу $x(t)$ має на меті представлення його у вигляді функціонального ряду – зваженої скінченної або безкінечної суми деяких складових:

$$x(t) = \sum_n c_n \psi_n(t), \quad (4.9)$$

де c_n – коефіцієнт розкладу, який відповідає вкладу відповідної складової у сигнал; ψ_n – елемент з деякого набору функцій. Для випадку вейвлетів цей набір не обов’язково повинен бути ортогональним базисом, а може також бути напівортогональним, неортогональним базисом або переповненим словником функцій.

В залежності від властивостей вейвлет-функцій ψ_n , на які розкладається сигнал, можна виділити такі класи вейвлет-перетворень:

- 1) Неперервне вейвлет-перетворення – виконується для неперервних та дискретних сигналів.
- 2) Розкладання сигналу по фреймам.
- 3) Ортогональні вейвлет-перетворення – клас перетворень сигналу, до якого належать: діадне ортогональне вейвлет-перетворення, вейвлет-перетворення з біортогональними базисами вейвлетів, розклад у вейвлет-пакети та вейвлети на інтервалі.

Розглянемо простір неперервних сигналів з обмеженою енергією $L^2(\mathbb{R})$. Якщо деяка функція $\psi(t) \in L^2(R)$, що задовольняє «умову допустимості»

$$C_\psi = \int_0^{+\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < +\infty \quad (4.10)$$

та має одиничну евклідову норму $\|\psi(t)\|=1$, то $\psi(t)$ може бути *материнським вейвлетом*.

Тут $\Psi(\omega)$ – спектр за Фур'є материнського вейвлета $\psi(t)$.

Задаючись параметром масштабу $a \in \mathbb{R}$, $a \neq 0$ та зсуву $\tau \in \mathbb{R}$, за правилом

$$\psi_{a,\tau}(t) \equiv |a|^{-\frac{1}{2}} \psi\left(\frac{t-\tau}{a}\right) \quad (4.11)$$

можна отримати набір вейвлет-функцій $\{\psi_{a,\tau}(t)\}$, що є масштабованими (розтягнутими або стиснутими залежно від величини коефіцієнту масштабування a) та зміщеними по аргументу (залежно від величини τ) копіями єдиної породжуючої материнської функції $\psi(t)$. Множник $\frac{1}{\sqrt{|a|}}$ носить технічний характер, він

використовується для того, щоб гарантувати рівність норми масштабованого вейвлета одиниці: $\|\psi_{a,\tau}(t)\|=1$. Сама материнська вейвлет-функція – також елемент цього набору:

$$\psi(t) = \psi_{1,0}(t).$$

Задачею вейвлет-аналізу сигналів є розклад аналізованого сигналу в суму масштабованих вейвлет-функцій (4.9). Значення відповідних коефіцієнтів розкладу визначаються як скалярні добутки сигналу на кожен масштабовану та зміщену вейвлет-функцію з набору $\{\psi_{a,\tau}(t)\}$:

$$\begin{aligned} W_\psi(a, \tau) &\equiv \langle x(t), \psi_{a,\tau}(t) \rangle = \\ &= \int_{-\infty}^{\infty} x(t) \psi_{a,\tau}^*(t) dt = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t-\tau}{a}\right) dt, \quad a, \tau \in \mathbb{R}, \quad a \neq 0, \end{aligned} \quad (4.12)$$

де $\psi_{a,\tau}^*$ – комплексно-спряжена материнська функція.

$W_\psi(a, \tau)$ – вейвлет-розклад сигналу $x(t)$.

Видно, що $W_\psi(a, \tau)$ є функцією двох змінних: масштабу та зміщення. Ця функція показує значення коефіцієнту (амплітуди), з яким в розклад (4.9) входить кожний вейвлет для всіх можливих значень a і τ . Говорять, що такий розклад відображає функцію однієї змінної $x(t) \in L^2(\mathbb{R})$ у функцію двох змінних $W_\psi(a, \tau) \in L^2(\mathbb{R} \times \mathbb{R})$.

Функція (4.12) є розкладом сигналу за всіма можливими стисненнями та зміщеннями однієї материнської функції.

Згідно з *теоремою Кальдерона, Гросмана і Морле*, якщо умова допустимості (4.10) виконується, то вейвлет-перетворення є неперервним та обмеженим, і існує обернене неперервне вейвлет перетворення. А отже, будь-яка функція $x(t) \in L^2(\mathbb{R})$ задовольняє рівностям:

$$\begin{aligned}
 x(t) &= \frac{1}{C_\psi} \int_0^{+\infty} \int_{R^+/0} \langle x(t), \psi_{a,\tau}(t) \rangle \psi_{a,\tau}(t) d\tau \frac{da}{a^2} = \\
 &= \frac{1}{C_\psi} \int_0^{+\infty} \int_{R^+/0} W_\psi(a, \tau) \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right) d\tau \frac{da}{a^2}.
 \end{aligned}
 \tag{4.13}$$

Цей вираз показує, що маючи коефіцієнти вейвлет-розкладу сигналу (4.12), можна відновити за ними початковий сигнал. Формула (4.13) називається *формулою Кальдерона* (названа іменем відомого американського математика Альберто П. Кальдерона).

Пара перетворень (4.12) та (4.13) називається відповідно *прямим та оберненим неперервним вейвлет-перетворенням (НВП) неперервних сигналів*.

Отже, *вейвлет-перетворення* – це розклад сигналу у набір функцій-вейвлетів, кожна з яких є розтягнутою або стисненою та зміщеною копією одної функції, яка називається материнським вейвлетом.

Можна довести, що НВП є перетворенням зі збереженням енергії, і для нього виконується аналог рівності Парсеваля:

$$\int_{-\infty}^{+\infty} |x(t)|^2 dt = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |W_\psi(a, \tau)|^2 d\tau \frac{da}{a^2}.
 \tag{4.14}$$

Якщо проаналізувати вираз для прямого НВП, то можна відмітити таке. Якщо параметр a (коефіцієнт масштабування) більший за одиницю та збільшується, то амплітуда і аргумент вейвлет-функції зменшується. Зменшення аргументу функції означає, що вона стає більш тривалою, повільною; в цьому випадку функція стає більш широкою в часовій області. Це, в свою чергу, приводить до того, що спектр вейвлет-функції стає більш низькочастотним. А, отже, коли буде розраховано інтеграл від добутку сигналу на такий низькочастотний вейвлет, в цьому коефіцієнті розкладу будуть відображені лише повільні, низькочастотні складові сигналу.

Навпаки, якщо масштабний коефіцієнт a буде меншим за одиницю, то аргумент масштабованої вейвлет-функції буде збільшуватися, її амплітуда також ставатиме більшою. В цьому випадку масштабований вейвлет стає стиснутим, коротшим в часовій області, а це еквівалентно тому, що його спектр буде високочастотним. Величина скалярного добутку сигналу на такий високочастотний вейвлет буде тим більшою, чим більш високочастотним є спектр ділянки сигналу.

Можна провести аналогію з тригонометричним базисом Фур'є: чим меншим є коефіцієнт перед змінною t в аргументі синуса, тим меншою є частота синусоїди (її спектр зсувається в область нижніх частот), і тим довшим є період (період займає більше часу); при збільшенні частоти, спектр гармоніки стає високочастотним, а період – коротшим. Якщо в сигналі є складові на певних частотах, то проекція сигналу на гармоніки цих частот буде великою.

Підбираючи параметр зміщення функції τ , можна переміщати масштабований вейвлет у часі. Нескінченні межі інтегрування показують, що коефіцієнти вейвлет-розкладу розраховуються для *всіх можливих комбінацій* значень коефіцієнта стиснення/розтягу, та зміщення вейвлет-функції у часі. Тобто, за допомогою

коефіцієнта масштабування a та параметру переносу τ можна створювати функції з високо- чи низькочастотним спектром та розміщувати їх в різні моменти часу. На такі складові і ведеться вейвлет-розклад сигналу. Отже, нестационарні сигнали з різними частотними компонентами, які розташовані на різних проміжках часу, можна описувати як суму різних вейвлет- функцій (4.9). Ілюстрація розрахунку вейвлет-перетворення наведена на рис. 4.6.

Можна сказати, що величина коефіцієнта розкладу відображає «схожість» ділянки сигналу за формою та спектральним складом на масштабований вейвлет. Ця схожість оцінюється локально, лише в околі того періоду часу, куди зараз зміщений вейвлет за допомогою параметра τ , і лише на ділянці, яка має таку ж тривалість, як і масштабований вейвлет. Ця унікальна властивість вейвлет-аналізу широко використовується в дослідженні сигналів різної природи.

Вклад певної складової вейвлет-розкладу до повної енергії сигналу для деякого масштабу a та зміщення τ може бути представлений на двовимірному графіку функції густини енергії для вейвлетів, яка називається *скейлограма*:

$$E(a, \tau) = |W_{\psi}(a, \tau)|^2.$$

Вираз (4.14) говорить про те, що оскільки вся енергія сигналу в часі дорівнює сумі енергій вейвлет-складових, то скейлограма може бути використана для дослідження розподілу енергій вейвлет-складових сигналу по масштабам та у часі. Приклади скейлограм подано на рис. 4.7. Колір на графіку відповідає абсолютному або відносному значенню енергії складової для кожного значення масштабу (вертикальна вісь) та зміщення (горизонтальна вісь).

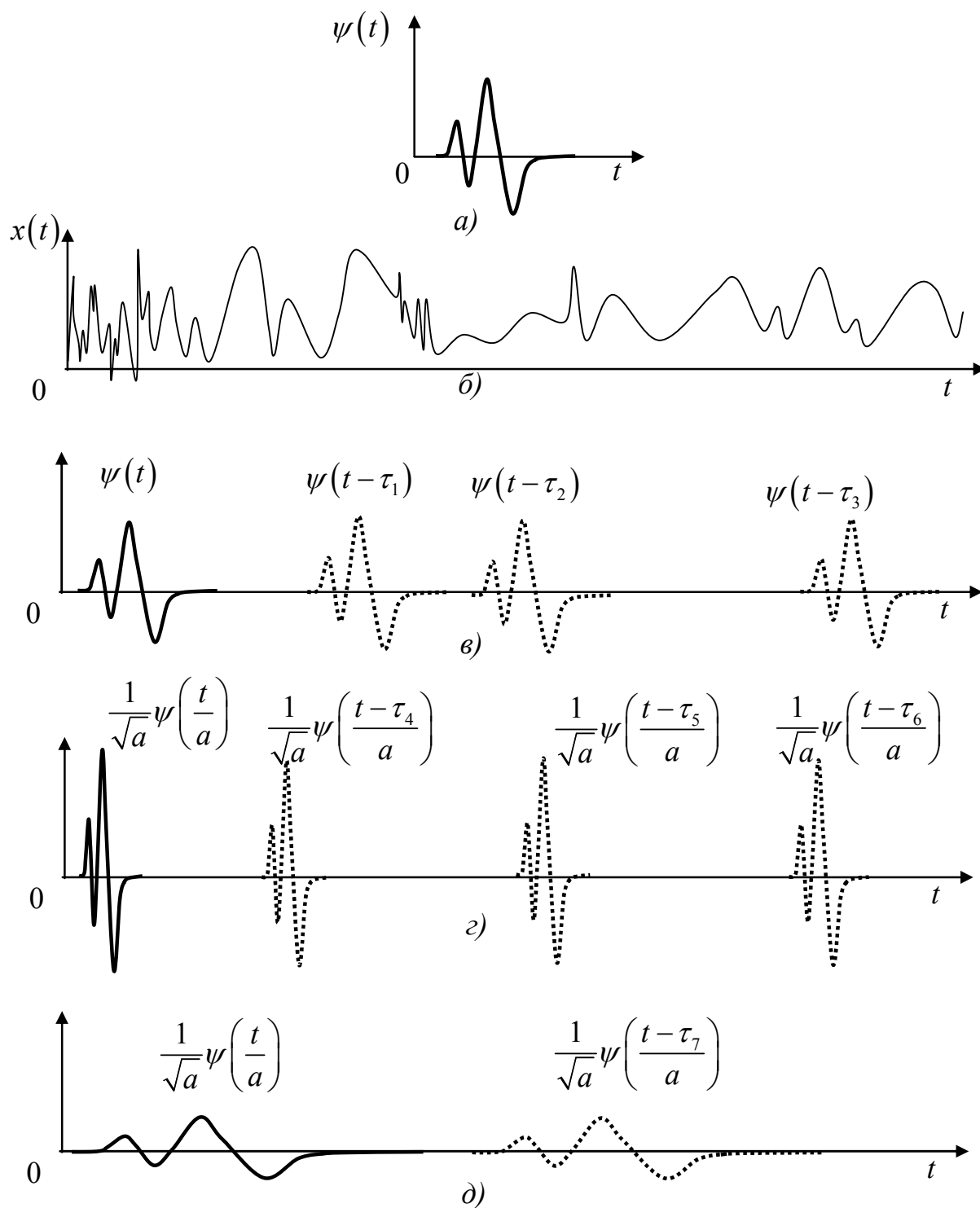
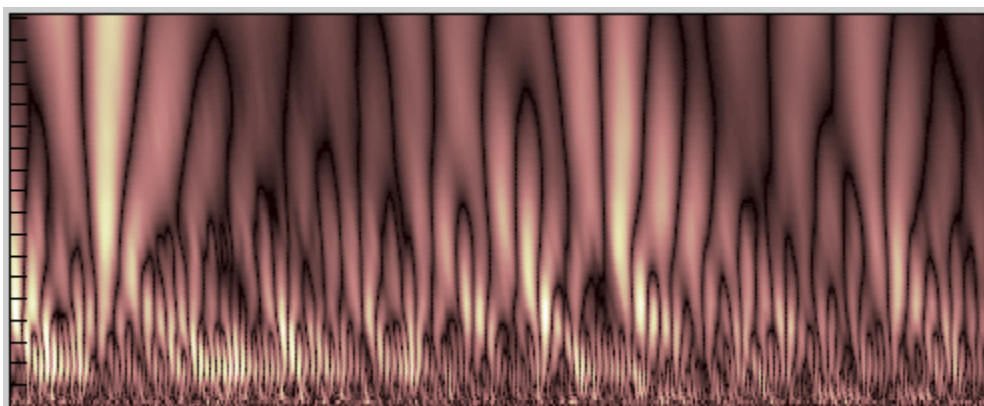
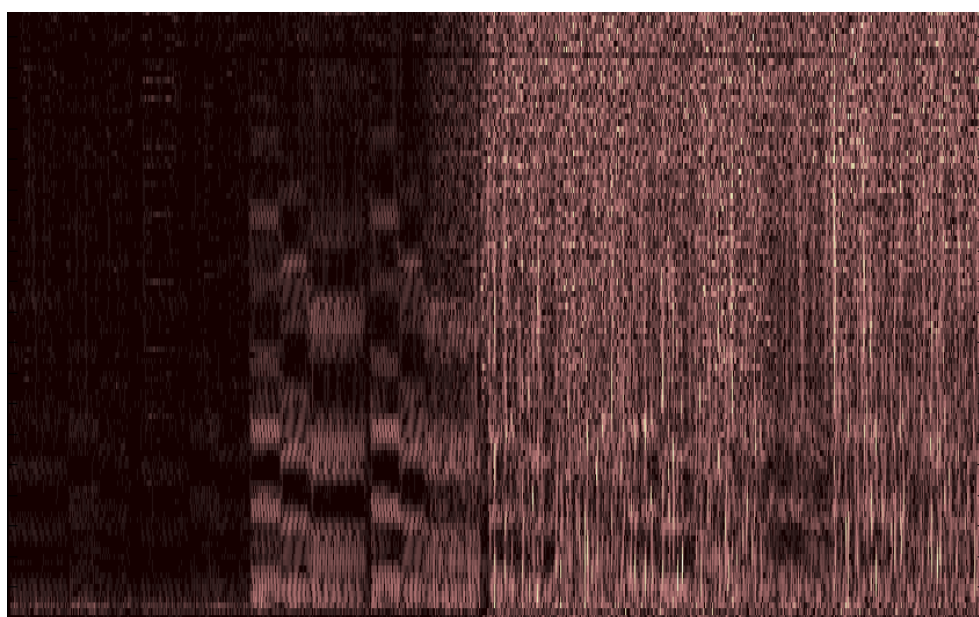


Рисунок 4.6 – а) материнська вейвлет-функція, б) сигнал, в) можливі зміщення немасштабованої материнської функції ($a = 1$), г) можливі зміщення при масштабах $0 < a < 1$, д) можливі зміщення при масштабах $a > 1$



а)



б)

Рисунок 4.7 – Скейлограма сигналу електроенцефалограми здорової людини, використано материнську функцію Морле (а), скейлограма запису пісні з рис. 4.2б, використано материнську функцію Добеші 4-го порядку (б)

Можна також досліджувати відносний вклад в енергію сигналу лише складових вейвлет-розкладу на певному масштабі a , та сформувати розподіл енергій за масштабом:

$$E(a) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} |W_\psi f(a, \tau)|^2 d\tau. \quad (4.15)$$

Піки в значеннях $E(a)$ відображають ті домінантні масштаби, вейвлет-складові на яких найбільш потужно представлені в сигналі. Така залежність втрачає інформацію про час появи в сигналі складових на певних масштабах. Аналогом виразу (4.15) для аналізу за Фур'є є розподіл потужності в спектрі сигналу.

Для того, щоб визначити, коли в сигналі присутні більш або менш потужні компоненти, можна проінтегрувати скейлограму по масштабам, отримавши в результаті функцію часу:

$$E(\tau) = \frac{1}{C_\psi} \int_{0+}^{+\infty} |W_\psi f(a, \tau)|^2 da. \quad (4.16)$$

Отримана функція вказує на розподіл енергії вейвлет-складових в часі. Більш докладний аналіз сигналу можна виконати, якщо розглядати у виразі (4.16) не всі масштаби, а лише певний проміжок. Таким чином можна дослідити, який вклад вейвлет-компонентів на певних масштабах у сигнал. Це дає змогу зв'язати часові та енергетичні характеристики сигналу.

Можна також перейти від масштабно-залежного вейвлет-спектру до частотно-залежного вейвлет-спектру, щоб порівнювати їх, та оцінювати розподіл потужності між частотними складовими на спектрограмі. Можна пов'язати масштаб розтягу/стиснення вейвлета з характеристичною частотою, з піковою частотою, центральною частотою. Але такий перехід є неоднозначним і досі немає остаточного методу співставлення частоти та масштабу вейвлета.

4.4. Материнські вейвлет-функції

Існують дві обов'язкові вимоги, яким повинна задовольняти материнська вейвлет-функція для того, щоб з її допомогою виконувалося пряме та обернене перетворення сигналу:

1) функція $\psi(t)$ повинна бути суттєво спадною на нескінченності:

$$|\psi(t)| < C(1 + |t|)^{-(1+\varepsilon)}, \quad (4.17)$$

де C, ε – деякі константи, $C, \varepsilon \in \mathbb{R}, \varepsilon > 0$;

2) функція $\psi(t)$ повинна бути гладкою. Гладкість можна оцінити за спадом спектральної функції материнського вейвлета:

$$|\Psi(\omega)| < C(1 + |\omega|)^{-(1+\varepsilon)}, \quad (4.18)$$

або за кількістю нульових моментів.

Умови (4.17) та (4.18) об'єднують, висуваючи до вейвлет-функції вимогу мати компактний носій у часовій та частотній областях, тобто якнайменшу замкнену область визначення, поза якою функція тотожно дорівнює нулю. Від величини носія залежить швидкість збіжності вейвлет-розкладу аналізованого сигналу за та концентрація енергії вейвлет-складових у частотно-часовому просторі.

Із виконання умови допустимості (4.10) випливає вимога

$$|\Psi(0)|^2 = 0, \quad (4.19)$$

тобто гармонічний спектр вейвлета повинен мати форму, схожу на амплітудно-частотну характеристику смугопропускаючого фільтра. Вираз (4.19) є еквівалентом виконання в часовій області умови

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0. \quad (4.20)$$

З (4.17) та (4.20) випливає, що материнська вейвлет-функція повинна мати коливний характер на деякому компактному носії, і мати нульове середнє значення. Отже, материнським вейвлетом може бути будь-яка функція, яка має нульове середнє значення, одиничну норму, та є «суттєво спадною в часі та частоті».

Існують два типи вейвлетів – які мають та не мають аналітичного виразу.

Історично першим ортогональним базисом вейвлетів був базис Хаара. Материнський вейвлет визначається за формулою:

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2}, \\ -1, & \frac{1}{2} \leq t < 1, \\ 0, & \text{в інших випадках} \end{cases}.$$

Іншим прикладом ортонормованого базису, за яким проводять розклад сигналів, є базис вейвлетів Літлвуда-Пелі з материнською функцією

$$\psi(t) = \frac{\sin(2\pi t) - \sin(\pi t)}{\pi t}.$$

Вейвлет Морле, який найбільш часто використовується із комплексних вейвлетів:

$$\psi(t) = \frac{1}{\sqrt[4]{\pi}} \left(e^{j\omega_0 t} - e^{-\frac{\omega_0}{2}} \right) e^{-\frac{t^2}{2}},$$

де ω_0 – «центральна частота» вейвлета, на якій спектр набуває максимального значення. Амплітудний спектр цього вейвлета: $H(\omega) = \sqrt{2\pi} e^{-\frac{(\omega-\omega_0)^2}{2}}$ – обмежена по частоті функція.

Вейвлет «мексиканський капелюх» – друга похідна функції Гаусса:

$$\psi(t) = \frac{2}{\sqrt{3}\sqrt[4]{\pi}} (1-t^2) e^{-\frac{t^2}{2}}.$$

Взагалі, всі похідні від функції Гаусса можуть бути материнськими вейвлетами. Материнський вейвлет Мейера задається не в часовій, а у частотній області:

$$\Psi(\omega) = \begin{cases} \frac{e^{\frac{i\omega}{2}}}{\sqrt{2\pi}} \sin \left[\frac{\pi}{2} \nu \left(\frac{3}{2\pi} |\omega| - 1 \right) \right], & \frac{2\pi}{3} \leq |\omega| < \frac{4\pi}{3}, \\ \frac{e^{\frac{i\omega}{2}}}{\sqrt{2\pi}} \cos \left[\frac{\pi}{2} \nu \left(\frac{3}{4\pi} |\omega| - 1 \right) \right], & \frac{4\pi}{3} \leq |\omega| < \frac{8\pi}{3}, \\ 0, & \text{в інших випадках,} \end{cases}$$

де $\nu(x)$ – функція, що задовольняє умовам:

$$\nu(x) = \begin{cases} 0, & x \leq 0, \\ 1, & x \geq 1, \end{cases}$$

$$\nu(x) + \nu(1-x) = 1.$$

Графіки деяких материнських вейвлетів наведено на рис. 4.8.

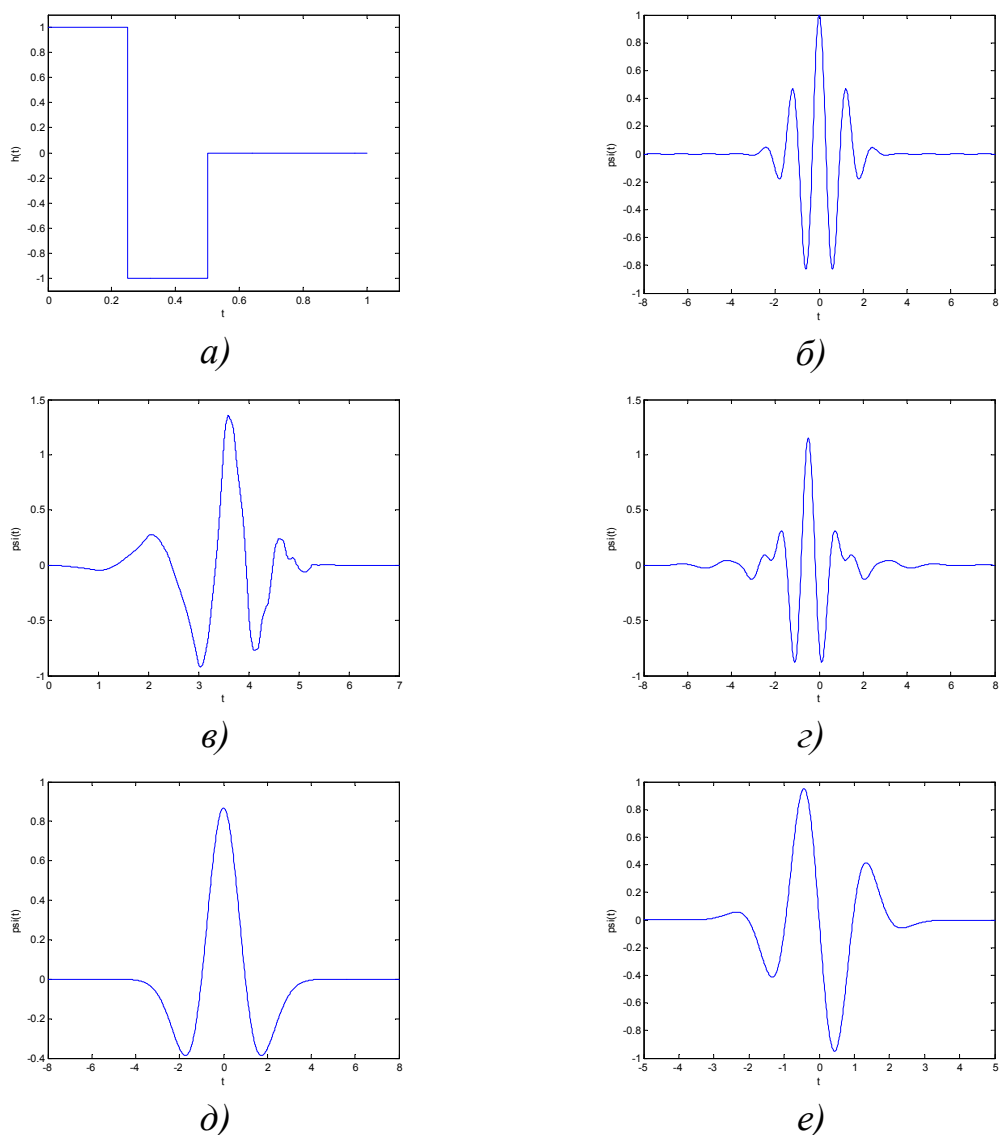


Рисунок 4.8 – Материнські вейвлет-функції: а) функція Хаара, б) вейвлет Морле, в) вейвлет Добеші 4-го порядку, г) вейвлет Мейера, д) вейвлет «мексиканський капелюх», е) вейвлет Гауса 5-го порядку

4.5. Розклад сигналів у фреймах

НВП сигналу (4.12) містить в загальному випадку нескінченну кількість коефіцієнтів розкладу, тому що для виконання формули Кальдерона (4.13) і коефіцієнт масштабування, і зміщення повинні приймати все можливі значення з множини дійсних чисел. На практиці таке виконати неможливо, тому потрібно знайти шлях до зменшення кількості коефіцієнтів, необхідних для точного відновлення досліджуваної функції за її вейвлет-розкладом.

У випадку НВП представлення сигналу $x(t)$ має нескінченну кількість складових ξ_n , які є елементами набору вейвлет-функцій (4.11) з нескінченною кількістю елементів. Існує можливість виконати представлення довільної функції $x(t) \in L^2(R)$ виду (4.13) з меншою кількістю складових частин за допомогою фреймів.

Побудувати обернене до (4.12) перетворення, яке б не потребувало врахування всіх коефіцієнтів a та τ , можна, якщо накласти на материнську вейвлет-функцію та на породжуваний нею набір додаткові обмеження. При цьому можна зменшити кількість коефіцієнтів, необхідних для відновлення сигналу. В цьому випадку отримується «дискретне» вейвлет-перетворення, при якому множина параметрів масштабу та зміщення є дискретною, а кількість елементів розкладу є меншою.

Для того, щоб дискретизувати параметри a та τ вейвлет-функції з метою зменшення кількості коефіцієнтів, обмежуються, по-перше, лише додатними значеннями масштабного коефіцієнта a (це вже було враховано в попередніх перетвореннях). Для дискретизації параметра масштабу a обирають

$$a = a_0^m, \quad m \in Z, \quad a_0 > 0, \quad a_0 \neq 1, \quad (4.21)$$

де a_0 – фіксований крок стиснення.

При такому виборі масштабів вони будуть змінюватися не довільно, а можуть бути лише рівними одному заздалегідь обраному числу в цілому ступені.

Для дискретизації параметра зміщення материнського вейвлета по часовій вісі також обирають значення, кратні деякому фіксованому кроку τ_0 ($\tau_0 > 0$). Для того, щоб при вейвлет-аналізі було враховано всі локальні особливості сигналу, масштабовану вейвлет-функцію треба в часі зміщувати так, щоб не було пропущено жодної ділянки сигналу. Можна показати, що для всіх значень $m \in Z$ ширина в часовій області функції $a_0^{-\frac{m}{2}} \psi(a_0^{-m}t)$ у a_0^m разів більша, ніж ширина немасштабованої функції $\psi(t)$. Тому в разі, коли параметр зміщення обирають як

$$\tau = n\tau_0 a_0^m, \quad (4.22)$$

дискретизовані вейвлети, отримані за (4.11) “накривають” при зміщеннях всю область, на якій заданий сигнал $x(t)$.

Вибір параметрів масштабу та зміщення (4.21) та (4.22) відповідає такій модифікації виразу для отримання набору вейвлет-функцій (4.11):

$$\psi_{a,\tau}(t) \equiv \psi_{m,n}(t) = a_0^{-\frac{m}{2}} \psi\left(\frac{t - n\tau_0 a_0^m}{a_0^m}\right) = a_0^{-\frac{m}{2}} \psi(a_0^{-m}t - n\tau_0). \quad (4.23)$$

Доведено, що функції, отримані за (4.23) можна використовувати для представлення довільного сигналу зі скінченною енергією лише у випадку, якщо вони складають фрейм.

Родина функцій $\{\varphi_j(t)\}_{j \in J}$ з гільбертового простору H називається *фреймом*, якщо існують такі константи $A > 0$, $B < \infty$, що для всіх функцій $x(t) \in H$ вірними є оцінки

$$A\|x(t)\|^2 \leq \sum_{j \in J} |\langle x(t), \varphi_j(t) \rangle|^2 \leq B\|x(t)\|^2. \quad (4.24)$$

Параметри A та B називають межами фрейма. Фрейм – це множина, в загальному випадку, залежних векторів, які можуть бути використані для розкладу кожного вектора даного простору у вигляді узагальненого представлення (4.9). Функції у фреймі не повинні бути ортогональні, зазвичай вони не є лінійно-незалежними, і коефіцієнти розкладу сигналу по фрейму можуть не бути однозначними. Якщо функції у фреймі лінійно-незалежні, то такий фрейм називається базисом Рітца. Перевагами використання фреймів є те, що накладені на них вимоги не такі жорсткі як у випадку ортогональних базисів, що дозволяє накладати на фрейми необхідні нам додаткові умови. Представлення сигналу за допомогою фрейму є найбільш “економічним” представленням серед представлень за будь-якими іншими наборами функцій.

Якщо межі фрейма рівні, то він називається жорстким. В цьому випадку виконується умова

$$\sum_{j \in J} |\langle x(t), \varphi_j(t) \rangle|^2 = A\|x(t)\|^2, \quad (4.25)$$

та вірною є формула відновлення функції за її проекціями на жорсткий фрейм:

$$x(t) = A^{-1} \sum_{j \in J} \langle x(t), \varphi_j(t) \rangle \varphi_j(t). \quad (4.26)$$

У випадку, коли всі елементи жорсткого фрейму мають одиничну норму і $A=1$, то фрейм є ортонормованим базисом.

В загальному випадку для фрейму $\{\varphi_j(t)\}_{j \in J}$ можна визначити дуальний йому фрейм $\{g_j(t)\}_{j \in J}$ з межами A^{-1} та B^{-1} :

$$A^{-1}\|x(t)\|^2 \leq \sum_{j \in J} |\langle x(t), g_j(t) \rangle|^2 \leq B^{-1}\|x(t)\|^2, \quad (4.27)$$

і можна побудувати формулу для відновлення значень сигналу за її проекціями на початковий фрейм:

$$x(t) = \sum_{j \in J} \langle x(t), \varphi_j(t) \rangle g_j(t). \quad (4.28)$$

Можна показати, що сигнал $x(t)$ можна відновити із розкладу за фреймом за допомогою чисельно стійкого алгоритму без явного розрахування дуального фрейму за ітеративною процедурою:

$$x_N(t) = x_{N-1}(t) + \frac{2}{A+B} \sum_{j \in J} [\langle x(t), \varphi_j(t) \rangle - \langle x_{N-1}(t), \varphi_j(t) \rangle] \varphi_j(t). \quad (4.29)$$

Отже, для можливості використання вейвлет-розкладу з дискретизованою множиною параметрів (4.21), (4.22) до родини вейвлет-функцій (4.23) висувається вимога, щоб вона складала фрейм.

Тому в разі суттєвої спадності у часовій та частотній областях та виконання умови (4.20) існує широкий спектр значень параметрів a_0 та τ_0 , для яких відповідні $\psi_{m,n}(t)$ утворюють фрейми. Від значень цих параметрів залежить швидкість збіжності процесу відновлення функції. В загальному випадку просторів кінцевої розмірності майже будь-яка множина векторів, що покриває весь простір, становить фрейм. В разі коли фрейм є жорстким, фрейм та дуальний фрейм є близькими, що спрощує алгоритм відновлення функції.

4.6. Діадне вейвлет-перетворення

Найбільш розповсюдженим є випадок, коли $a_0 = 2$, а $\tau_0 = 1$, тобто $\tau = na$. В цьому випадку родина вейвлет-функцій буде отримана за виразом:

$$\psi_{m,n}(t) = 2^{-\frac{m}{2}} \psi(2^{-m}t - n), \quad m, n \in \mathbb{Z}. \quad (4.30)$$

Говорять, що отримані так вейвлети задані на *діадній сітці*, тобто крок зміни значень масштабу є степенем двійки, а зміщення в часі кратне двом.

Вейвлет-перетворення сигналу для такого вибору коефіцієнтів називається *діадним вейвлет-перетворенням* (ДВП).

В переважній кількості випадків при використанні дискретної діадної сітки вимагають, щоб вейвлети, отримані за (4.30), були ортонормовані. Вимога ортонормованості – це те, чим ми платимо за можливість використовувати для представлення сигналу не всю нескінченну кількість коефіцієнтів розкладу, а лише обмежену дискретну множину:

$$\int_{-\infty}^{+\infty} \psi_{m,n}(t) \psi_{m',n'}^*(t) dt = \begin{cases} 1, & m = m', \quad n = n' \\ 0, & m \neq m', \quad n \neq n' \end{cases}$$

Ортогональність складових розкладу (вейвлет-функцій) говорить про те, що інформація, яку кожна з функцій «витягує» з сигналу, не повторюється деінде в іншому коефіцієнті, а зберігається лише у значенні єдиного вейвлет-коефіцієнту. А отже, представлення сигналу в такому ортогональному базисі не містить надлишковості, бо комбінація ортонормованих векторів може повністю визначити сигнал.

В цьому разі замість неперервного вейвлет-перетворення як неперервної функції двох аргументів (масштабу та зміщення), маємо скінченну кількість значень:

$$D_{m,n} \equiv \langle x(t), \psi_{m,n}(t) \rangle = \int_{-\infty}^{+\infty} x(t) \psi_{m,n}(t) dt. \quad (4.31)$$

Обернене перетворення має загальний вигляд:

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} D_{m,n} \cdot \psi_{m,n}(t). \quad (4.32)$$

Зауважимо, що значення (4.31) є коефіцієнтами розкладу сигналу $x(t)$ в по системі функцій (4.30). Вираз (4.32) є узагальненим рядом Фур'є сигналу $x(t)$ відносно цієї системи. Отже, щоб ці вирази були вірними, набір масштабованих вейвлетів, за якими ведеться розклад, повинен складати ортогональний базис.

Формули (4.31) – (4.32) визначають *пряме і обернене ортогональне діадне вейвлет-перетворення неперервного сигналу $x(t)$* . Іноді його не зовсім коректно називають «дискретним вейвлет-перетворенням».

Представлення сигналу (4.32) має чітку ієрархічну структуру. При фіксованому значенні m (рівень розрізнення, resolution) масштаб вейвлетів не змінюється, а отже всі вейвлети для кожного рівня розрізнення є зсунутими копіями один одного. При збільшенні розрізнення на одиницю, вдвічі стискаються вейвлети та вдвічі зменшується величина зсуву по часовій вісі (діадність). Таким чином на кожному масштабі покриваються вся область часу, на якій заданий аналізований сигнал.

Треба зауважити, що поняття «дискретне» або «діадне» у вейвлет-перетворенні відноситься лише до *способу вибору коефіцієнтів масштабування*, які тепер не будуть змінюватися неперервно, а повинні набувати лише певні значення. Можна говорити про як про НВП, так і про ДВП неперервних або дискретних сигналів.

Якщо сигнал неперервний, то для нього можна виконати НВП (в цьому випадку коефіцієнти масштабування та зміщення можуть бути довільними), або ДВП (при цьому буде діадна зміна масштабів та зміщень).

Якщо сигнал дискретний, то для проведення його НВП коефіцієнти масштабування все одно можуть набувати будь-які значення. Зміщення також може бути довільним. Але треба у виразах для прямого та оберненого НВП замінити інтеграли сумами та зробити так, щоб вирази справджувались при підстановці в них дискретних відліків замість неперервної функції. Також треба розробити методику отримання дискретних відліків масштабованих материнських функцій, та забезпечити при цьому співпадіння сітки вузлів для дискретного сигналу та вейвлета, а також одиничну норму. Також треба переконатися, що при зміщенні вейвлета сітка вузлів, на якій він заданий, зберігається, тобто зміщати лише на цілу кількість кроків дискретизації. Ці перетворення не є тривіальними, але їх можливо виконати.

Для виконання ДВП дискретного сигналу все вищесказане також є вірним, але і масштаби, і зміщення є діадними.

4.7. Кореляційний аналіз детермінованих сигналів

Часто перед дослідниками виникають такі задачі:

– пошук в сигналі ділянок, які є схожими на певні, заздалегідь визначені зразки (патерни). Це задача, наприклад, виявлення в електроенцефалограмі коливань, які відповідають передвісникам епілепсій, порівняння надісланого та віддзеркаленого від літака імпульсів та т.і.;

– визначення, наскільки два сигнали схожі один на інший за формою. Для таких сигналів треба оцінити, чи пов'язані між собою зміни форми сигналів, чи синхронно відбуваються в сигналах коливання, чи однакові параметри сигналів;

– визначити, чи в одному і тому самому сигналі присутні ділянки, які схожі одна на одну, які повторюються через деякий час, або знайти, чи є в сигналі періодичності і чому дорівнює період сигналу.

Одним зі шляхів вирішення такого роду задач є *кореляційний аналіз сигналів*. В цьому розділі увага буде зосереджена лише на аналізі *детермінованих* сигналів. В теорії випадкових сигналів також звертаються до питань кореляції, (кореляційний або спектрально-кореляційний аналіз), але їм буде присвячений окремий розділ.

Нехай маємо два скінченні сигнали з обмеженою енергією $x_1(t)$ та $x_2(t)$ (та $x_1[n]$ і $x_2[n]$). Як кількісний показник того, наскільки вони схожі один на одного, можна обрати інтеграл від їх добутку. Якщо при переході від одного моменту часу до іншого сигнали змінюються схоже, то їх поточкові добутки будуть великими, і інтеграл буде мати велике значення. Якщо значення одного і іншого сигналу змінюються випадково та не пов'язані одні з іншими, то ці поточкові добутки будуть невеликими, мати різні знаки, та значення інтегралу буде близьким до нуля. Для того, щоб значення запропонованого показника не залежало від тривалості, і показувало лише на те, синхронними чи випадковими є зміни в сигналах, інтеграл треба також поділити на довжину сигналів.

Також треба врахувати, що між змінами в сигналах, навіть синхронними, може бути певна затримка. Наприклад, спочатку зростають значення одного сигналу, а потім через деякий час так само зростають значення іншого. Щоб врахувати це в показнику схожості, його треба розраховувати для випадків, коли один сигнал є зафіксованим у часі, а інший – зсувається в тому чи іншому напрямку відносно першого. І для кожного значення зсуву треба розраховувати інтеграл від добутків миттєвих значень сигналів. Отже, показник схожості двох сигналів буде не числом, а функцією від параметра зміщення сигналів один відносно одного. Ця функція буде характеризувати кореляцію між двома сигналами при різних запізненнях одного відносно іншого.

Взаємнокореляційна функція (ВКФ) для скінченних неперервних та дискретних сигналів буде визначатися за виразами:

$$r_{12}(\tau) = \frac{1}{T} \int_0^T x_1(t) x_2(t - \tau) dt, \quad (4.33)$$

$$r_{12}[j] = \frac{1}{N} \sum_{n=0}^{N-1} x_1[n] x_2[n - j], \quad (4.34)$$

де τ та j – параметри, які показують зміщення сигналів один відносно одного.

Видно, що вирази (4.33) та (4.34) будуть залежати від значень сигналів (амплітуд), і тому вони не завжди будуть відображати саме схожість форми сигналів. Щоб уникнути цього, треба нормувати функцію на енергії сигналів. В результаті

отримаємо вирази, які можна використовувати при практичних розрахунках. Остаточно отримаємо такі модифіковані вирази для ВКФ:

$$B_{12}(\tau) = \frac{r_{12}(\tau)}{\frac{1}{T} \sqrt{\int_0^T x_1^2(t) dt \cdot \int_0^T x_2^2(t) dt}}, \quad (4.35)$$

$$B_{12}[j] = \frac{r_{12}[j]}{\frac{1}{N} \sqrt{\sum_{n=0}^{N-1} x_1^2[n] \cdot \sum_{n=0}^{N-1} x_2^2[n]}}. \quad (4.36)$$

ВКФ показує схожість між двома сигналами. Значення функцій $B_{12}(\tau)$ та $B_{12}[j]$ завжди будуть лежати між -1 та +1. -1 відповідає випадку, коли зростання одного сигналу приводить до зменшення іншого, а +1 – коли сигнали змінюються абсолютно однаково. Якщо значення ВКФ дорівнює нулю, це може значити, що зміни сигналів або не пов'язані, або механізм зв'язку між сигналами не можна виявити за допомогою ВКФ.

Значення ВКФ при $\tau = 0$ та $j = 0$ називають *коефіцієнтами (взаємної) кореляції* двох сигналів.

Якщо треба визначити, чи схожим є сигнал на свою зміщену копію (наприклад, чи є в ньому ділянки, які повторюються, чи періодичний сигнал), то можна скористатися таким же підходом, як і для ВКФ, але брати лише один сигнал. В такому випадку розраховують *автокореляційну функцію* (кореляційну функцію) (АКФ):

$$r_x(\tau) = \frac{1}{T} \int_0^T x(t)x(t-\tau)dt, \quad (4.37)$$

$$r_x[j] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]x[n-j]. \quad (4.38)$$

Ці вирази також можна модифікувати аналогічно до (4.35) та (4.36):

$$B_x(\tau) = \frac{\int_0^T x(t)x(t-\tau)dt}{\int_0^T x^2(t)dt}, \quad (4.39)$$

$$B_x[j] = \frac{\sum_{n=0}^{N-1} x[n]x[n-j]}{\sum_{n=0}^{N-1} x^2[n]}. \quad (4.40)$$

Треба також зважати на особливості розрахунків ВКФ та АКФ для скінченних сигналів. Оскільки в розрахунках використовують сигнали скінченної тривалості, то при достатньо великих зміщеннях τ та j величина області перекриття двох сигналів буде невеликою. Значення двох сигналів вже не будуть формувати парні

добутки. Тому і значення ВКФ для цих аргументів може бути меншим, ніж насправді, внаслідок крайових ефектів.

Властивості ВКФ та АКФ:

1. Значення ВКФ при нульовому зміщенні дорівнює потужності сигналу (для періодичних сигналів – середній потужності за період):

$$r_x(0) = P.$$

Якби усереднення за період в виразах було відсутнім, то значення при $\tau = 0$ було б рівним енергії сигналу:

$$r(0) = \int_{-\infty}^{+\infty} x^2(t) dt = E.$$

2. АКФ є парною функцією:

$$r_x(-\tau) = r_x(\tau).$$

3. Для ВКФ виконується властивість:

$$r_{12}(-\tau) = r_{21}(\tau),$$

тобто зміна знаку еквівалентна перестановці сигналів місцями.

4. Значення АКФ при нульовому зміщенні є максимально можливим значенням. Зі збільшенням зміщення сигналів АКФ затухає.

5. Максимальне значення ВКФ може бути при будь-якому значенні взаємного зміщення двох сигналів.

6. АКФ періодичного сигналу є періодичною функцією. ВКФ сигналів з однаковим періодом є періодичною функцією з таким самим періодом. Період ВКФ сигналів з різними періодами дорівнює меншому періоду з двох сигналів.

Розглянемо зв'язок між ВКФ та спектральними характеристиками сигналів. Нехай сигнали $x_1(t)$ та $x_2(t)$ мають спектри відповідно $S_1(\omega)$ та $S_2(\omega)$. Для загального випадку нескінченних сигналів скористаємось таким виразом для ВКФ:

$$r'_{12}(\tau) = \int_{-\infty}^{+\infty} x_1(t) x_2(t - \tau) dt.$$

Знайдемо пряме інтегральне перетворення Фур'є для $r'_{12}(\tau)$:

$$\begin{aligned} \int_{-\infty}^{+\infty} r'_{12}(\tau) e^{-j\omega\tau} d\tau &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} x_1(t) x_2(t - \tau) dt \right) e^{-j\omega\tau} d\tau = \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} x_1(t) x_2(t - \tau) dt \right) e^{-j\omega\tau} e^{j\omega(t-\tau)} d\tau = \int_{-\infty}^{+\infty} x_1(t) e^{-j\omega t} dt \int_{-\infty}^{+\infty} x_2(t - \tau) e^{j\omega(t-\tau)} d\tau = \\ &= S_1(\omega) S_2^*(\omega). \end{aligned}$$

Отже, видно, що спектр ВКФ є добутком спектральних характеристики сигналів, одна з яких є комплексно спряженою. Важливим наслідком цього є те, що якщо спектри двох сигналів займають різні області частот (не перекриваються), то ВКФ таких сигналів рівна нулю для всіх можливих зсувів. Отже, сигнали з такими спектрами принципово не можуть бути корельованими. Але треба зважати на те, що в реальному житті навіть такі сигнали можуть змінюватися синхронно, і залежати

один від одного. Для виявлення зв'язку між такими сигналами треба використовувати інші характеристики (міри синхронності, інформаційну місткість сигналів, ентропійні та фрактальні характеристики та т.і.)

Якщо виконати аналогічні перетворення для АКФ, то можна отримати:

$$\int_{-\infty}^{+\infty} r'_x(\tau) e^{-j\omega\tau} d\tau = S_1(\omega) S_2^*(\omega) = |S_1(\omega)|^2.$$

Отже, спектр АКФ дорівнює спектру потужності сигналу. Також видно, що оскільки для спектру АКФ береться квадрат спектру сигналу, то він не буде залежати від фазового спектру. Отже, АКФ не залежить від фазо-частотних характеристик спектру сигналу.

4.8. Перетворення Хартлі

Перетворення Фур'є відображає дійсні сигнали в комплексну область, а також є несиметричним по j (при переході від прямого до оберненого перетворення відбувається зміна знаку). Воно вимагає проведення розрахунків з комплексними числами. При його реалізації треба в загальному випадку зберігати окремо дійсну та уявну частини спектрів, що потребує додаткових ресурсів.

Увага дослідників була направлена на позбавлення від цих незручностей, але водночас на збереження переваг гармонічного аналізу.

У 1942 році Ральф Хартлі запропонував перетворення сигналів, яке не залучає комплексних чисел. Це перетворення виникло внаслідок потреби аналізувати чисто дійсні, а не комплексні послідовності одновимірних та двовимірних даних. Обробку таких даних можна було здійснювати в області дійсних чисел з допомогою взаємно симетричного прямого та оберненого перетворень.

Було введено пару формул для *прямого та оберненого перетворення Хартлі* сигналу $x(t)$:

$$\begin{aligned} H(f) &= \int_{-\infty}^{+\infty} x(t) \text{cas}(2\pi ft) dt, \\ x(t) &= \int_{-\infty}^{+\infty} H(f) \text{cas}(2\pi ft) df, \end{aligned} \tag{4.41}$$

де $\text{cas}(t) = \cos t + \sin t$ (рис. 4.5).

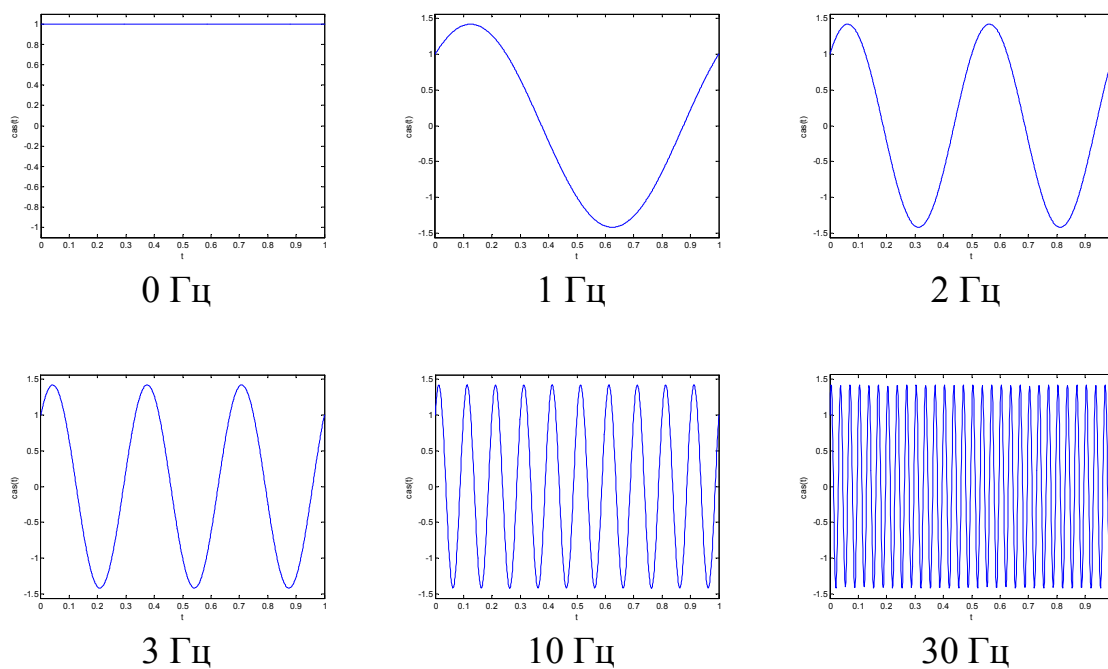


Рисунок 4.9 – Приклади базисних функцій базису Хартлі

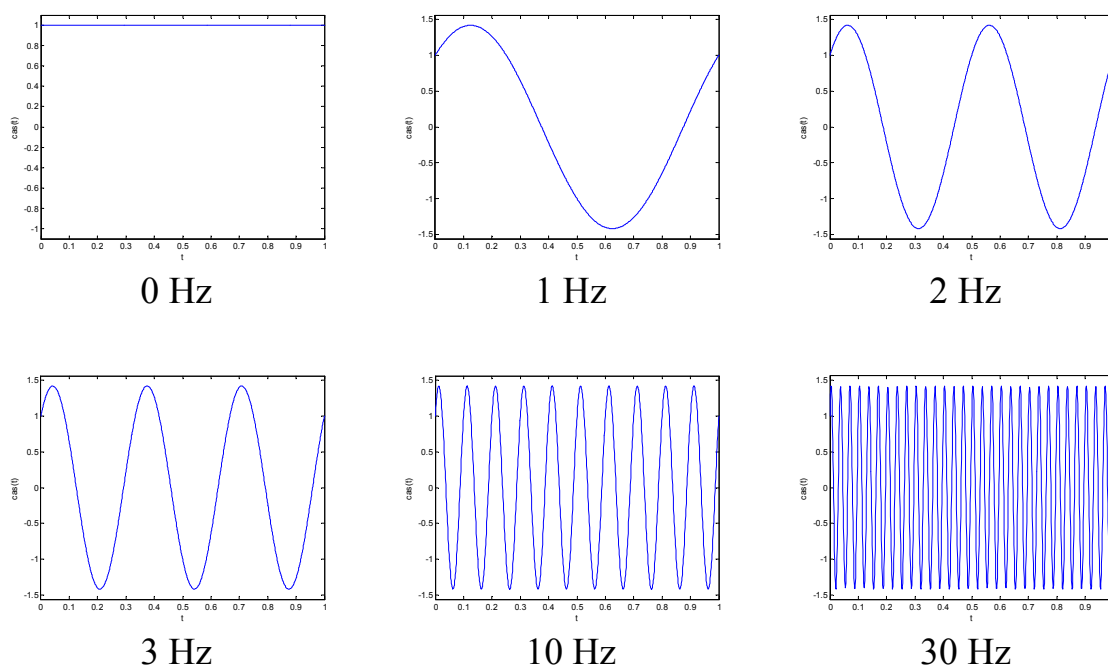


Рисунок 4.9 – Приклади базисних функцій базису Хартлі

Для знаходження зв'язку між перетворенням Хартлі та перетворенням Фур'є, вводять до розгляду парну та непарну компоненти спектру Хартлі $H(f)$:

$$E(f) = \frac{H(f) + H(-f)}{2} = \int_{-\infty}^{+\infty} x(t) \cos(2\pi ft) dt,$$

$$O(f) = \frac{H(f) - H(-f)}{2} = \int_{-\infty}^{+\infty} x(t) \sin(2\pi ft) dt.$$

Ці вирази також відомі як косинусне та синусне перетворення Фур'є.

Комплексний спектр сигналу за Фур'є буде дорівнювати

$$F(f) = E(f) - jO(f).$$

В свою чергу, перетворення Хартлі можна знайти як різницю між дійсною та уявною частиною перетворення Фур'є:

$$H(f) = \operatorname{Re}(F(f)) - \operatorname{Im}(F(f)).$$

Енергетичний спектр за Фур'є сигналу можна розрахувати

$$P(f) = \frac{(H^2(f) + H^2(-f))}{2}$$

Фазовий спектр сигналу за допомогою перетворення Хартлі визначається як

$$\varphi(f) = \operatorname{arctg} \frac{-O(f)}{E(f)}$$

У 1983 році Рональд Брейсуелл запропонував *пряме та обернене перетворення Хартлі для дискретних сигналів*:

$$H[m] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \operatorname{cas} \left[\frac{2\pi nm}{N} \right],$$

$$x[n] = \sum_{m=0}^{N-1} H[m] \operatorname{cas} \left[\frac{2\pi nm}{N} \right].$$

Перетворення Хартлі використовується в обробці дійсних сигналів, аналізі двовимірних та тривимірних зображень, проектуванні швидких перетворень, інтерполяції сигналів та т.і.

4.9. Перетворення Уолша-Адамара

Розглянемо ще одне перетворення, яке останніми роками набуло дуже широкого вжитку в системах цифрового зв'язку.

Матрицею Адамара H_N називається ортогональна квадратна матриця порядку N ($N = 2^l$, $l = 0, 1, 2, 3, \dots$), елементами якої є дійсні числа ± 1 . Найпростішою матрицею Адамара є матриця другого порядку:

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Для побудови матриць Адамара вищих порядків використовується така теорема:
Якщо H_N – матриця Адамара порядку N , то матриця:

$$H_{2N} = \begin{bmatrix} H_N & H_N \\ H_N & -H_N \end{bmatrix}$$

є матрицею Адамара порядку $2N$.

Отже, можна рекурсивно будувати матриці Адамара різних порядків, маючи матриці попередніх порядків. Матрицю Адамара четвертого порядку можна отримати, маючи матрицю другого порядку:

$$H_4 = \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \end{bmatrix},$$

а матриця Адамара восьмого порядку отримується з чотирьох матриць четвертого порядку:

$$H_8 = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \end{bmatrix} & \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \end{bmatrix} \\ \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} & \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \end{bmatrix} & \begin{bmatrix} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} & \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \\ \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} & \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \end{bmatrix} \end{bmatrix}.$$

Для того, щоб отримати елемент матриці Адамара довільного порядку одразу, не записуючи матриці рекурсивно, можна використати таке правило. Пронумеруємо рядки та стовпці матриці Адамара двійковими числами від 0 до $N-1$. Наприклад, для матриці порядку $N = 2^3 = 8$ це будуть числа 000, 001, 010, 011, 100, 101, 110, 111. Якщо необхідно визначити, який елемент розташований на перетині рядка з номером k та стовпця з номером n , то треба розглядати двійкові номери рядка та стовпця як вектори, та скористатися формулою:

$$had(k, n) = (-1)^{\langle n, k \rangle}, \quad (4.42)$$

де $\langle k, n \rangle$ – скалярний добуток «векторів» k та n .

Наприклад, елемент матриці H_8 на перетині 3-го рядка та 5 стовпця буде визначатися:

$$had(3, 5) = (-1)^{\langle [0,1,0], [100] \rangle} = (-1)^{0 \cdot 1 + 1 \cdot 0 + 0 \cdot 0} = (-1)^0 = 1.$$

Зауважимо ще раз, що нумерація починається з нуля.

Рядки матриці Адамара

$$had(k, n), \quad k = 0 \dots N-1$$

можна розглядати як *функції дискретної змінної n* , визначені в цілочислених точках $0, 1, 2, \dots, N-1$ інтервалу від 0 до $N-1$. Ці функції називаються *функціями Уолша*. Змінна k визначає номер функції (номер рядка в матриці Адамара), змінна n – дискретний час або номер відліку. Ці функції належать до N -вимірного простору сигналів.

Наприклад, для матриці Адамара восьмого порядку матимемо функції:

$$had(000, n) = [1, 1, 1, 1, 1, 1, 1, 1]$$

$$had(001, n) = [1, -1, 1, -1, 1, -1, 1, -1]$$

$$had(010, n) = [1, 1, -1, -1, 1, 1, -1, -1]$$

$$had(011, n) = [1, -1, -1, 1, 1, -1, -1, 1]$$

$$had(100, n) = [1, 1, 1, 1, -1, -1, -1, -1]$$

$$had(101, n) = [1, -1, 1, -1, -1, 1, -1, 1]$$

$$had(110, n) = [1, 1, -1, -1, -1, -1, 1, 1]$$

$$had(111, n) = [1, -1, -1, 1, -1, 1, 1, -1]$$

Графіки функцій Уолша 16-го порядку наведені на рис. 4.10.

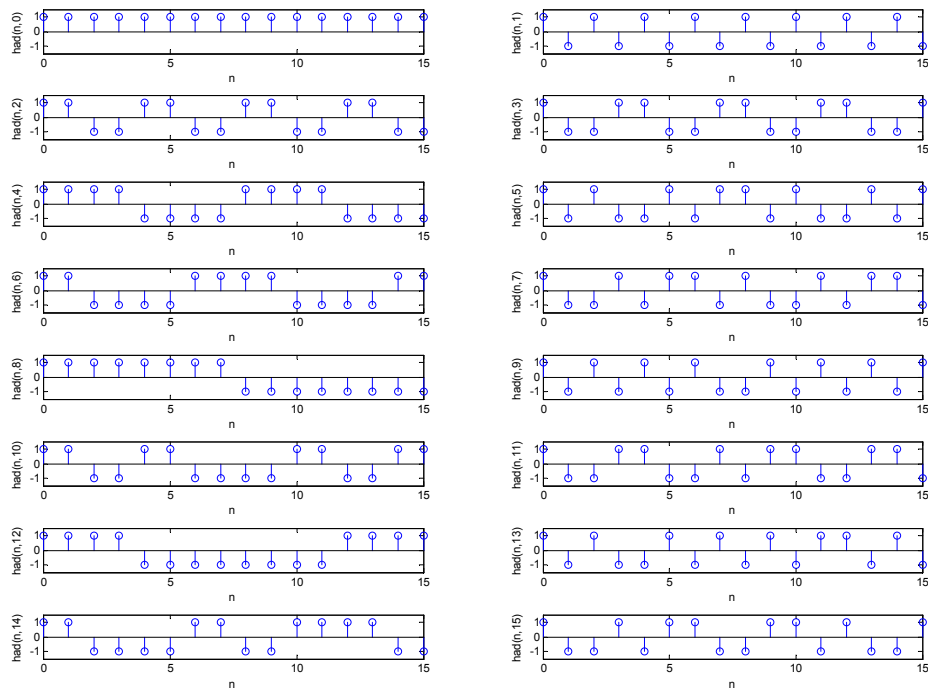


Рисунок 4.10 – Функції Уолша порядку $N = 16$

Функції Уолша були відкриті та використані для аналізу сигналів Джозефом Леонардом Уолшем в 1923 році, хоча в матричній формі були відомі ще задовго до цього таким вченим як Сильвестр та Адамар.

Можна показати, що функції Уолша є ортогональними:

$$\begin{aligned}\langle had(k_1, n), had(k_2, n) \rangle &= \sum_{n=0}^{N-1} had(k_1, n) \cdot had(k_2, n) = \\ &= \begin{cases} \neq 0, & k_1 \neq k_2, \\ 0, & k_1 = k_2. \end{cases}\end{aligned}$$

та утворюють повну систему функцій. Тому можна будь-який сигнал $x[n]$ представити у вигляді розкладу за цією системою функцій. Коефіцієнти розкладу будуть визначатися за виразом:

$$W[k] = \frac{\langle x[n], had(k, n) \rangle}{\|had(k, n)\|^2}.$$

Норма функції Уолша дорівнює:

$$\begin{aligned}\|had(k, n)\| &= \sqrt{\langle had(k_1, n), had(k_2, n) \rangle} = \sqrt{\sum_{n=0}^{N-1} had(k, n) \cdot had(k, n)} = \\ &= \sqrt{\sum_{n=0}^{N-1} (-1)^{\langle n, k \rangle} \cdot (-1)^{\langle n, k \rangle}} = \sqrt{\sum_{n=0}^{N-1} (-1)^{2\langle n, k \rangle}} = \sqrt{\sum_{n=0}^{N-1} 1} = \sqrt{N}.\end{aligned}$$

Пряме перетворення Уолша дискретних сигналів:

$$W[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] had(k, n). \quad (4.43)$$

Обернене перетворення Уолша:

$$x[n] = \sum_{k=0}^{N-1} W[k] had(k, n). \quad (4.44)$$

На практиці використовується також і інша нумерація функцій, що відповідає іншому способу впорядкування узагальнених гармонік (складових розкладу за Уолшем). Найбільш широко вживаними є впорядкування за Пелі та за Уолшем. Система Пелі отримується двійковою інверсією номеру функції за Адамаром і подальшою перестановкою функцій, а система Уолша розташовує функції за зростанням кількості змін знаку функції на інтервалі (по «частотам» слідування).

Перевагою використання функцій Уолша для розкладу сигналів порівняно з спектральним розкладом за Фур'є є те, що розрахунок спектру Уолша зводиться до додавання відліків сигналу. На відміну від перетворення Фур'є, аналіз сигналів за Уолшем не потребує множення комплексних чисел, на що потрібно більше часу, а для збереження дійсної та уявної частин потрібно достатньо велика кількість розрядів. Розрахунок коефіцієнта розкладу (проекції сигналу на базисну функцію) є множенням відліків сигналу на 1 або -1 та подальшим додаванням з урахуванням знаку, що особливо зручно робиться у швидкодіючих цифрових системах.

Перетворення Уолша та функції Уолша використовують у завадостійкому кодуванні сигналів, аналізу та синтезу пристроїв логічної дії, дослідженні логічних функцій, в стандарті цифрового зв'язку CDMA.

Варто зазначити, що функції Уолша, як і функції Фур'є, враховують поведінку сигналу одразу на всьому проміжку існування, тобто не мають властивості часової локалізації. Як і функції Хаара, функції Уолша є кусково-неперервними та містять розриви, тому вони є оптимальними для представлення імпульсоподібних сигналів з крутими фронтами. Також функції Уолша корисно застосовувати для аналізу двовимірних сигналів з порушеннями неперервності, наприклад, зображень в астрономії та спектроскопії.

4.10. Перетворення Гілберта

Нехай маємо дійсний сигнал $x(t)$. Розглянемо деякий комплексний сигнал $\tilde{x}(t)$, що побудований на його основі, і для якого сигнал $x(t)$ буде дійсною частиною:

$$\tilde{x}(t) = x(t) + jx_{\perp}(t). \quad (4.45)$$

Сигнал $\tilde{x}(t)$ назвемо *аналітичним сигналом* для нашого заданого сигналу $x(t)$. Тут сигнал $x_{\perp}(t)$ є результатом дії деякого оператора на початковий сигнал; тобто сигнали $x(t)$ та $x_{\perp}(t)$ повинні бути пов'язані деяким перетворенням. Нижче ми обґрунтуємо вимоги до цього перетворення та задамо його.

Для комплексного аналітичного сигналу, заданого за (4.45), можна задати поняття *огибаючої* як модуля цього сигналу:

$$X(t) = |\tilde{x}(t)| = \sqrt{\operatorname{Re}^2(\tilde{x}(t)) + \operatorname{Im}^2(\tilde{x}(t))} = \sqrt{x^2(t) + x_{\perp}^2(t)}, \quad (4.46)$$

а також *фази* як аргументу аналітичного сигналу:

$$\Phi(t) = \arg(\tilde{x}(t)) = \arctg \frac{\operatorname{Im}(\tilde{x}(t))}{\operatorname{Re}(\tilde{x}(t))} = \arctg \frac{x_{\perp}(t)}{x(t)} \quad (4.47)$$

Після визначення огибаючої та фази можна записати аналітичний сигнал в експоненціальній формі:

$$\tilde{x}(t) = X(t)e^{j\Phi(t)}.$$

Оскільки фаза цього сигналу нам відома, то можна записати частоту аналітичного сигналу. За визначенням, частота є швидкістю зміни фази, отже, вона дорівнює похідній від фази по часу:

$$\omega(t) = \frac{d\Phi(t)}{dt}.$$

До цього часу ми не зважали на те, як саме буде отримано уявну частину аналітичного сигналу $x_{\perp}(t)$. Але від способу її отримання буде залежати те, який

саме смисл ми зможемо вкласти в поняття «модуля» та «фази» аналітичного сигналу. Взагалі, існує багато можливостей задати цю уявну частину, оскільки ми маємо один початковий сигнал $x(t)$, а фактично хочемо задати на його основі два сигнали – огибаючу і фазу. Але ми оберемо такий спосіб, щоб отриманий аналітичний сигнал мав сенс з точки зору сигналів електронних систем, струмів та напруг. Висунемо такі вимоги до способу отримання уявної частини аналітичного сигналу:

1. Малі зміни в початковому сигналі $x(t)$ повинні приводити до малих змін в огибаючій та фазі аналітичного сигналу. Для цього оператор, за допомогою якого отриманий $x_{\perp}(t)$, повинен бути неперервним.

2. Із уявлень про сигнали в технічних системах відомо, що фаза не є енергетичним параметром, отже фаза, визначена за (4.47), не повинна залежати від енергії сигналу. При збільшенні або зменшенні енергії сигналу фаза повинна залишатися такою самою. Фаза – це арктангенс відношення результату дії оператора на сигнал до самого цього сигналу:

$$\Phi(t) = \arctg \frac{x_{\perp}(t)}{x(t)} = \arctg \frac{H(x(t))}{x(t)}.$$

Множення сигналу на константу (що еквівалентно збільшенні його енергії) не повинне міняти фазу сигналу, тобто цю константу повинно бути можна винести за знак оператора:

$$\frac{H(\lambda x(t))}{\lambda x(t)} = \frac{\lambda H(x(t))}{\lambda x(t)} = \frac{H(x(t))}{x(t)},$$

Тобто оператор, який задає уявну частину сигналу, повинен бути однорідним.

3. Задамо сигнал виду $x(t) = U_m \cos(\omega_0 t + \varphi)$. В ньому явно видно, що огибаюча $X(t) = U_m$, а фаза $\Phi(t) = \omega_0 t + \varphi$. Отже, перетворення, за яким буде отримано уявну частину аналітичного сигналу, повинне приводити до таких саме значень огибаючої і фази.

Можна довести, що єдине перетворення, яке задовольняє наведеним трьома вимогам – це *перетворення Гілберта*. Отже, уявна частина аналітичного сигналу $x_{\perp}(t)$ визначається як результат перетворення Гілберта для сигналу $x(t)$, а саме:

$$H(x(t)) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau. \quad (4.48)$$

Видно, що перетворення Гілберта – лінійне перетворення, внаслідок лінійності інтеграла. При розгляді (4.48) видно, що цей вираз є інтегралом згортки функції $x(t)$ та функції $\frac{1}{\pi t}$:

$$H(x(t)) = x_{\perp}(t) = x(t) * \frac{1}{\pi t}.$$

Отже, внаслідок властивостей згортки, в частотній області перетворення Гілберта буде являти собою добуток спектральної густини сигналів $x(t)$ та $\frac{1}{\pi t}$:

$$F_{x_{\perp}}(j\omega) = F_x(j\omega) \cdot F_H(j\omega).$$

Знайдемо спектр функції $\frac{1}{\pi t}$ за Фур'є:

$$\begin{aligned} F_H(j\omega) &= \int_{-\infty}^{+\infty} \frac{1}{\pi t} e^{-j\omega t} dt = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{t} (\cos \omega t - j \sin \omega t) dt = \\ &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{t} (\cos \omega t - j \sin \omega t) dt = \\ &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{t} \cos \omega t dt - \frac{j}{\pi} \int_{-\infty}^{+\infty} \frac{1}{t} \sin \omega t dt. \end{aligned}$$

Перший із двох інтегралів в сумі є інтегралом від добутку непарної функції $\frac{1}{t}$ на парну функцію косинуса, який береться в симетричних межах, отже, він буде дорівнювати нулю. Другий інтеграл є табличним, отже

$$F_H(j\omega) = -\frac{j}{\pi} \pi \cdot \text{sign}(\omega) = -j \text{sign}(\omega),$$

де $\text{sign}(\omega)$ – функція знака:

$$\text{sign}(\omega) = \begin{cases} 1, & \omega \geq 0 \\ -1, & \omega < 0. \end{cases}$$

Знайдена спектральна функція $F_H(j\omega) = -j \text{sign}(\omega)$ є комплексною частотною характеристикою деякої лінійної системи, яка при подачі на вхід сигналу $x(t)$ дає на виході сигнал $x_{\perp}(t)$ (рис. 4.11). Ця система буде виконувати пряме перетворення Гілберта для сигналу, поданого на її вхід.

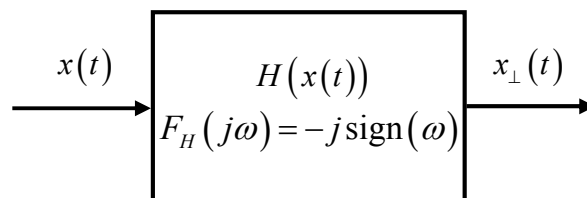


Рисунок 4.11 – Система, яка реалізує пряме перетворення Гілберта

З теорії лінійних стаціонарних систем відомо, що комплексна частотна характеристика та імпульсна характеристика системи пов'язані перетворенням Фур'є. Як видно з виразу (4.48), за яким в часовій області отримується вихідний

сигнал, імпульсна характеристика системи, яка реалізує перетворення Гілберта дорівнює

$$h_H(t) = \frac{1}{\pi t}.$$

Знайдемо АЧХ та ФЧХ системи з рис. 4.6. Отримаємо модуль функції $F_H(j\omega) = -j \operatorname{sign}(\omega)$:

$$|F_H(j\omega)| = |-j \operatorname{sign}(\omega)| = \sqrt{-j \operatorname{sign}(\omega) \cdot j \operatorname{sign}(\omega)} = \sqrt{1} = 1,$$

оскільки квадрат знакової функції завжди буде рівний 1.

Аргумент $F_H(j\omega)$:

$$\arg F_H(j\omega) = \operatorname{arctg} \frac{-\operatorname{sign}(\omega)}{0} = \begin{cases} \frac{\pi}{2}, & \omega < 0 \\ -\frac{\pi}{2}, & \omega \geq 0, \end{cases}$$

Отже, система, яка реалізує перетворення Гілберта (її ще називають *фільтром Гілберта*) не змінює амплітуду вхідного сигналу (оскільки АЧХ дорівнює 1), і повертає фазу всіх спектральних складових вхідного сигналу на $-\frac{\pi}{2}$ для додатних

частот і на $\frac{\pi}{2}$ для від'ємних частот (рис. 4.12). оскільки амплітуда сигналу при проходженні через фільтр Гілберта не змінюється, то незмінною залишається також норма і енергія сигналу. Іншими словами, оператор перетворення Гілберта є унітарним оператором.

Запишемо знову рівняння $F_{x_\perp}(j\omega) = F_x(j\omega) \cdot F_H(j\omega)$, та підставимо в нього щойно отриманий спектр $F_H(j\omega)$:

$$F_{x_\perp}(j\omega) = F_x(j\omega)(-j \operatorname{sign}(\omega)). \quad (4.49)$$

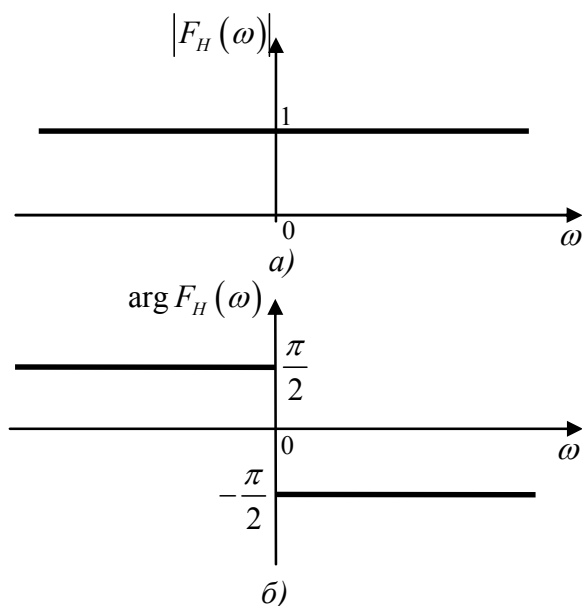


Рисунок 4.12 – Спектральні характеристики фільтра Гілберта: а) АЧХ, б) ФЧХ

Попередній вираз – це запис прямого перетворення Гілберта в частотній області: спектр сигналу $x_{\perp}(t)$ отримується множенням спектру початкового сигналу $x(t)$ на функцію $-j\text{sign}(\omega)$.

Для знаходження оберненого перетворення Гілберта в частотній області треба виразити спектр початкового сигналу $x(t)$ через спектр сигналу $x_{\perp}(t)$. Помножимо ліві і праву частини (4.49) на $j\text{sign}(\omega)$:

$$j\text{sign}(\omega)F_{x_{\perp}}(j\omega) = F_x(j\omega)(-j\text{sign}(\omega))j\text{sign}(\omega),$$

звідки

$$F_x(j\omega) = j\text{sign}(\omega)F_{x_{\perp}}(j\omega). \quad (4.50)$$

Пара виразів (4.49) та (4.50) є прямим та оберненим перетвореннями Гілберта, записаними в частотній області. Видно, що пряме перетворення відрізняється від оберненого лише знаком. З врахуванням цього можна одразу записати обернене перетворення Гілберта в часовій області:

$$H(x_{\perp}(t)) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x_{\perp}(\tau)}{t - \tau} d\tau. \quad (4.51)$$

Важливою властивістю перетворення Гілберта є те, що сигнал і його перетворення Гілберта є ортогональними. Доведемо це, розрахувавши скалярний добуток:

$$\langle x(t), x_{\perp}(t) \rangle = \int_{-\infty}^{+\infty} x(t)x_{\perp}^*(t) dt.$$

Запишемо цей скалярний добуток в частотній області, скориставшись узагальненою рівністю Парсеваля щодо рівності скалярного добутку в часовій та частотній областях:

$$\begin{aligned}
\langle x(t), x_{\perp}(t) \rangle &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_x(j\omega) F_{x_{\perp}}(j\omega) d\omega = \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_x(j\omega) (j \operatorname{sign} \omega) F_x^*(j\omega) d\omega = \\
&= \frac{j}{2\pi} \int_{-\infty}^{+\infty} |F_x(j\omega)|^2 \operatorname{sign} \omega d\omega = 0.
\end{aligned}$$

Остання рівність дорівнює нулю, оскільки квадрат модуля спектру сигналу є парною функцією, яка множиться на непарну функцію знаку, а інтеграл береться в симетричних межах. Отже, сигнал та його перетворення Гілберта є ортогональними. Знайдемо спектральну характеристику аналітичного сигналу. Нехай сигнал $x(t)$ має вузькосмуговий спектр, який сконцентрований в околі деякої частоти ω_0 (рис. 4.13)

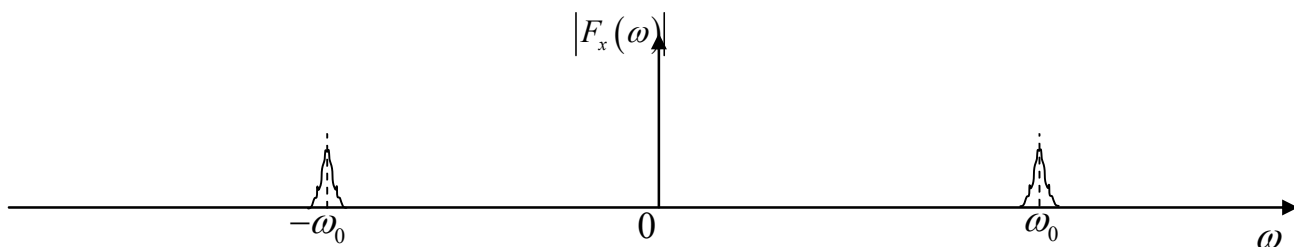


Рисунок 4.13 – Спектр вузькосмугового сигналу $x(t)$

Аналітичний сигнал (4.45) рівний сумі сигналу $x(t)$ та його перетворення по Гілберту $x_{\perp}(t)$. Внаслідок лінійності перетворення Фур'є, спектр аналітичного сигналу буде рівний:

$$\begin{aligned}
F_x(j\omega) &= F_x(j\omega) + jF_{x_{\perp}}(j\omega) = \\
&= F_x(j\omega) + j(-j \operatorname{sign} \omega) F_x(j\omega) = \\
&= F_x(j\omega)(1 + \operatorname{sign} \omega) = \begin{cases} 2F_x(j\omega), & \omega \geq 0 \\ 0, & \omega < 0 \end{cases}
\end{aligned}$$

Отже, спектр аналітичного сигналу буде присутній лише на додатних частотах, по формі він буде таким, як і спектр початкового сигналу $x(t)$, але амплітуда кожної гармоніки буде вдвічі більшою (рис. 4.14).

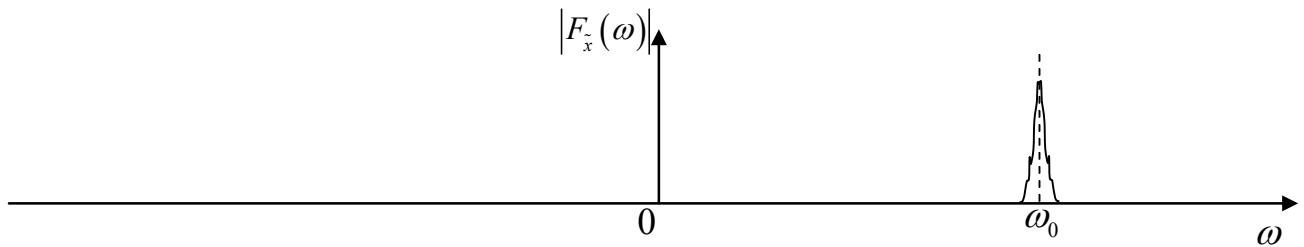


Рисунок 4.14 – Спектр аналітичного сигналу для сигналу $x(t)$

Зробимо зауваження щодо поняття огинаючої $X(t)$. Зміст цієї назви для отриманої за (4.46) функції можна пояснити та обґрунтувати таким:

- із (4.46) видно, що $X(t) \geq x(t)$, тобто графік $X(t)$ не перетинає графік початкового сигналу $x(t)$;
- розглянемо прямі, які дотичні до графіків $x(t)$ і $X(t)$ в точках, де ці графіки торкаються один одного. Для цього знайдемо похідну огинаючої:

$$\begin{aligned}
 X'(t) &= \frac{d}{dt} \left(\sqrt{x^2(t) + x_{\perp}^2(t)} \right) = \\
 &= \frac{1}{2 \sqrt{x^2(t) + x_{\perp}^2(t)}} (2x(t)x'(t) + 2x_{\perp}(t)x'_{\perp}(t)) = \\
 &= \frac{x(t)x'(t) + x_{\perp}(t)x'_{\perp}(t)}{X(t)} \Rightarrow \\
 X(t)X'(t) &= x(t)x'(t) + x_{\perp}(t)x'_{\perp}(t).
 \end{aligned}$$

Останній вираз пов'язує огинаючу, початковий сигнал та його перетворення Гілберта і їх похідні. Розглянемо випадок, коли $X(t) = x(t)$, тобто коли графіки огинаючої і сигналу торкаються один одного. При цьому із (4.46) слідує, що $x_{\perp}(t) = 0$, і значить, похідні огинаючої та сигналу рівні між собою. Згадуючи, що похідна функції в певній точці має смисл кута нахилу дотичної до графіка функції в цій точці, можна сказати, що початковий сигнал і його огинаюча мають в точках торкання спільні дотичні.

Введемо поняття *комплексної огинаючої*. Запишемо аналітичний сигнал в експоненціальній формі з використанням (4.46) та (4.47), та представимо фазу цього сигналу у вигляді $\Phi(t) = \omega_0 t + \phi(t)$:

$$\begin{aligned}\tilde{x}(t) &= x(t) + jx_{\perp}(t) = X(t)e^{j\Phi(t)} = X(t)e^{j(\omega_0 t + \phi(t))} = \\ &= X(t)e^{j\phi(t)}e^{j\omega_0 t} = \dot{S}(t)e^{j\omega_0 t}.\end{aligned}$$

Це суто математична підстановка, яка може бути зроблена незважаючи на те, який саме є дійсний закон зміни фази аналітичного сигналу. $\dot{S}(t)$ називається комплексною огибаючою. Попереднє рівняння можна інтерпретувати як те, що аналітичний сигнал дорівнює деякому комплексному сигналу $\dot{S}(t)$, який в часовій області помножений на комплексний експоненціальний сигнал $e^{j\omega_0 t}$. Згідно з властивостями перетворення Фур'є можна сказати, що таке множення в часовій області приведе до того, що спектр сигналу $\dot{S}(t)$ в частотній області буде зміщений по вісі частот на ω_0 вліво (рис. 4.15).

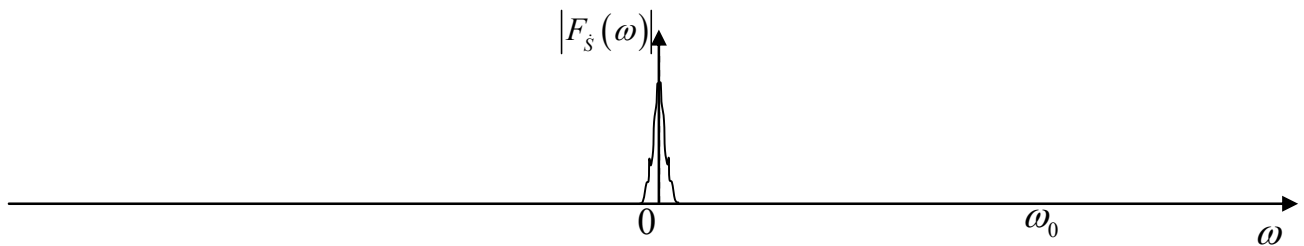


Рисунок 4.15 – Спектр комплексної огибаючої $\dot{S}(t)$ аналітичного сигналу

Іншими словами, аналітичний сигнал можна розглядати як деякий сигнал $\dot{S}(t)$, що мав спектр на нижніх частотах в околі нульової частоти, але який був помножений в часовій області на $e^{j\omega_0 t}$, і таким чином його спектр був зміщений в окіл частоти ω_0 . Іншими словами, спектр комплексної огибаючої є спектром аналітичного сигналу, який зміщений на частоту ω_0 . В загальному випадку спектр комплексної огибаючої не буде симетричним відносно нульової частоти, оскільки при математичних викладках для його отримання не були висунуті ніякі вимоги до частоти ω_0 . Зокрема, якщо ми аналізуємо деякий вузькосмуговий сигнал $x(t)$ зі спектром, сконцентрованим в околі деякої частоти ω_0 , то вибір частоти, на яку «зміщений» спектр комплексної огибаючої, є природнім. Для такого сигналу можна ввести поняття аналітичного сигналу $\tilde{x}(t)$ та комплексної огибаючої $\dot{S}(t)$. Тоді можна сказати, що наш початковий дійсний сигнал зв'язаний з комплексною огибаючою:

$$x(t) = \text{Re}(\dot{S}(t)e^{j\omega_0 t}).$$

Розглянемо випадок, коли аналізуємо деякий довільний сигнал $x(t)$, для якого отримали комплексну огибающую. Нехай $\dot{S}(t) = u(t) + jv(t)$, тоді наш сигнал $x(t)$ можна виразити через дійсну і уявну частини $\dot{S}(t)$:

$$\begin{aligned} x(t) &= \operatorname{Re}\left((u(t) + jv(t))e^{j\omega_0 t}\right) = \operatorname{Re}\left((u(t) + jv(t))(\cos \omega_0 t + j \sin \omega_0 t)\right) = \\ &= \operatorname{Re}\left(u(t)\cos \omega_0 t + u(t)j \sin \omega_0 t + jv(t)\cos \omega_0 t + jv(t)j \sin \omega_0 t\right) = \\ &= u(t)\cos \omega_0 t - v(t)\sin \omega_0 t. \end{aligned}$$

В результаті ми отримали фундаментальну властивість: довільний дійсний сигнал $x(t)$ можна представити у вигляді суми двох синусоїдальних коливань одної частоти (зміщених по фазі на $\frac{\pi}{2}$ один відносно одного), амплітуди кожного з яких змінюються в часі за деякими законами. Сигнали $u(t)$ та $v(t)$ називаються *синфазною* та *квадратурною* складовими для сигналу $x(t)$. Ще раз наголосимо, що частота ω_0 в попередньому виразі теоретично може обиратися довільно, оскільки від неї не залежить отримання комплексної огибающей $\dot{S}(t)$. На практиці необхідно зважати на можливість появи накладення спектрів аналітичного сигналу і комплексної огибающей. Також треба мати на увазі, що поняття комплексної огибающей має сенс лише при задаванні частоти ω_0 .

Застосування цієї властивості і перетворення Гілберта в теорії сигналів дуже широке. Воно застосовується там, де потрібна генерація або описання комплексних сигналів в часовій області, зокрема в таких застосуваннях:

- квадратурна модуляція та демодуляція, що використовується, зокрема, в системах цифрового зв'язку. При цьому на одній несівній частоті ω_0 в одному каналі зв'язку передається одночасно два сигнали $u(t)$ та $v(t)$, тобто досягається ущільнення каналу;
- аналіз сигналів в радарх і сонарах, аналіз дво- і тривимірних сигналів на основі запису їх у вигляді комплексних сигналів;
- дослідження проходження сигналів через лінійні системи. Метод заміни сигналів на їх комплексні огибаючі з метою дослідження їх проходження через лінійні системи називається «методом низькочастотного еквівалента».

4.11. Підсумки до розділу 4 та рекомендована література

В цьому розділі висвітлено розширення класичного аналізу за Фур'є, а саме спектрально-часового аналізу, який дозволяє спостерігати за зміною спектрального складу сигналу в часі. Він є продовженням та розширенням віконної обробки, і має

використовуватися скрізь, де сигнали змінюють свої властивості на інтервалі аналізу.

Фундаментальним недоліком спектрального аналізу є те, що складові розкладу сигналу при ньому вважаються такими, що існують протягом всього інтервалу аналізу. Щоб уникнути цього, введено поняття розкладу сигналу на складові різної тривалості. Найпершим прикладом такого підходу до аналізу було перетворення Хаара, але найбільш розповсюдженим та потужним методом аналізу сигналів зі змінними характеристиками є вейвлет-перетворення.

Вейвлет-перетворення є розкладом на складові різної тривалості, кожна з яких має свої характеристики. Тому залежно від того, яку материнську вейвлет-функцію обрано для отримання розкладу, такі характеристики аналізованого сигналу і будуть виявлені. Розглянуто два види аналізу неперервних сигналів: неперервне та діадне вейвлет-перетворення, які відрізняються способами обрання масштабів.

Розглянуто основи кореляційного аналізу детермінованих сигналів, який треба використовувати для оцінки поведінки сигналу в часі. Для визначення періодичностей чи схожості двох сигналів введені авто- та взаємнокореляційні функції та коефіцієнти кореляції.

Також зроблено огляд інших, відмінних від спектрального за Фур'є та вейвлетного аналізу. Показано, що сигнали можна розкладати в різних базисах (Хартлі, Хаара, Уолша), і залежно від того, якими базисними функціями користуватися, буде отримано те чи інше представлення сигналу. Оскільки загальний підхід до розкладу сигналів в базисі один і той самий, володіння ним дозволяє досліднику, обираючи відповідний базис, вивчати ті чи інші властивості одного і того самого сигналу.

Також розглянуто перетворення Гілберта, при якому для заданого дійсного сигналу формується аналітичний сигнал, а потім визначається його комплексна огинаюча. Зокрема, було показано, що для даного сигналу можна отримати його синфазну і квадратурну складову для певної несівної частоти.

Питання спектрально-часового аналізу добре висвітлені в [2, 7, 22]. Докладно про вейвлет-аналіз можна дізнатись з підручників [20, 25-27], в цих же книгах описане перетворення Хаара як один з видів вейвлет-перетворення. Більш глибоке математичне обґрунтування вейвлет-аналізу можна вивчати за роботами [14, 28-29]. Перетворення Уолша-Адамара докладно описане в [12], перетворенню Хартлі присвячена книга [30]. Кореляційний аналіз детермінованих сигналів описаний в [7].

4.12. Питання для самоперевірки

1. Для чого і як роблять спектрально-часовий аналіз сигналів за Фур'є? Що зміниться в результаті при різних тривалостях вікон та різних кроках зміщення вікна?
2. В чому особливість перетворення Хаара порівняно зі спектрально-часовим аналізом? Як отримують базисні функції Хаара?

3. В чому сутність вейвлет-перетворення, за яким виразом отримують вейвлет-розклад сигналу та скейлограму? Що слідує з формули Кальдерона щодо можливості відновлення сигналу із його вейвлет-розкладу?
4. Що таке материнська функція вейвлет-перетворення, які вони бувають і як з них отримуються масштабовані вейвлети?
5. В чому полягає розклад сигналу у фреймах та діадне вейвлет-перетворення?
6. Як і для чого проводять кореляційний аналіз сигналів? Які коефіцієнти і які функції можна отримати в результаті кореляційного аналізу?
7. Що таке перетворення Хартлі, чим воно відрізняється від перетворення Фур'є?
8. Що таке перетворення Уолша-Адамара, як отримують базисні функції для перетворення?
9. Що таке аналітичний сигнал і як отримати його фазу і огибаючу?
10. Яким є АЧХ та ФЧХ системи, яка реалізує перетворення Гілберта?
11. Як отримати квадратурне представлення для довільного сигналу?

5. ОСНОВИ ФІЛЬТРАЦІЇ НЕПЕРЕРВНИХ ТА ДИСКРЕТНИХ СИГНАЛІВ

Equation Chapter 1 Section 5

5.1. Шуми та перешкоди в сигналах

Всі сигнали, які треба аналізувати, вимірюються в середовищі лабораторії, кабінету або в польових умовах. Це означає, що поряд з тим об'єктом, який ми досліджуємо, і сигнали якого вимірюємо, існують також інші об'єкти, які можуть генерувати деякі сигнали. Якщо ці сторонні сигнали раптом потраплять на вхід нашої вимірювальної системи, то вони долучаться до корисного сигналу від об'єкта спостереження. Отже, в результуючому сигналі, який буде виміряно та передано для обробки та аналізу, будуть присутні дві складові: корисна (від об'єкту спостереження) та некорисна (іноді її називають шумовою). Шумова складова може заважати аналізу властивостей корисного сигналу, тому треба її позбутися перед цим.

Шум – будь-який небажаний сигнал, який заважає вимірюванню, передачі, обробці корисного сигналу. Результатом змішування в одному вимірювальному каналі корисного сигналу та шуму є спотворення сигналу – небажана зміна його форми, параметрів, властивостей.

Види шумів в сигналах:

- шуми електродів, якими проводиться вимірювання сигналів. Особливо актуальними є такі шуми при дослідженні біологічних об'єктів. Власні шуми електродів – це зміна їх вихідного сигналу, яка відбувається через особливості взаємодії матеріалу електрода із навколишнім середовищем. При контакті електрода зі шкірою людини можна виділити такі шуми: *тепловий шум* в активних опорах ланцюга електрод-електроліт; *дробовий шум* внаслідок подолання носіями струму потенціального бар'єру електрод-електроліт, *шум дрейфу* власної ЕРС електрода, *надлишковий шум*, який пов'язаний із нестабільністю подвійного електричного шару на поверхні електрода та корозією електрода;

- власні шуми електронної апаратури, які накладаються на корисний сигнал (шуми транзисторів, трансформаторів тощо);

- мережева перешкода 50 Гц, яка проникає в корисний сигнал, що вимірюється в будівлях, обладнаних мережею промислового електроживлення.

Також можуть бути випадки, коли джерело шуму знаходиться в самому об'єкті і генерує додатковий сигнал, що вимірюється одночасно із корисним сигналом.

Наприклад, при вимірюванні сигналів електроенцефалограми (низькоамплітудна низькочастотна електрична активність мозку) часто на вхід реєстратора потрапляє також і сигнал електрокардіограми (електричної активності серця). Він є набагато більшим за амплітудою, тому якщо обидва ці сигнали потрапили на вхід каналу підсилювача біопотенціалів, ЕЕГ буде складно розрізнити на фоні ЕКГ. В даному випадку ЕКГ є шумовим сигналом (перешкодою) для ЕЕГ,

якого треба позбутися. Якщо вдасться вилучити із сигналу ЕЕГ непотрібний в даному випадку сигнал ЕКГ, то аналіз ЕЕГ буде виконаний більш якісно.

Варто пам'ятати, що шум є також сигналом, який несе інформацію про джерело шуму. Шуми та спотворення є головними факторами, які обмежують роботу вимірювальних систем та систем передачі даних скрізь, де сигнал не можна ізолювати від впливу шумів: медичні застосування, телекомунікації, розпізнавання голосу та зображень. Тому моделювання та розробка методів видалення шумів та спотворень є одною з головних задач теорії сигналів. Від можливості описати та промоделювати шумовий процес залежить успішність використання того чи іншого методу обробки сигналів. Ми повинні використати знання про характеристики шумів для того, щоб боротися з шумами.

Розглянемо ситуацію, коли в лабораторії вимірюється електрокардіограма людини. Зовнішній вигляд одного кардіокомплекса та його спектр подано на рис. 5.1а. Видно, що в часовій області можна побачити всі властивості ЕКГ, зубці, інтервали та сегменти. Оскільки в лабораторії присутня мережа побутового електроживлення (50 Гц, 220 В), то навколо пацієнта та електрокардіографа існує змінне електромагнітне поле. Вигляд синусоїдальної напруги та її спектр подано на рисунку 5.1б. В ідеальному випадку спектр такої синусоїдальної напруги складається з одної спектральної лінії на частоті 50 Гц, але в реальності частота мережі електроживлення змінюється в деяких межах, а форма напруги не ідеальна. Тому маємо певне розширення спектру сигналу мережевої перешкоди.

Змінне поле від мережі буде діяти на електроди, які накладені на пацієнта, тому в сигнал буде проникати шумова складова – синусоїдальне коливання частотою 50 Гц. Вона буде спотворювати форму корисного сигналу, і аналізувати його буде складно. Приклад сигналу, який можна отримати в результаті змішування синусоїдального сигналу та електрокардіограми, подано на рисунку 5.1в. Видно, що в такому сигналі дуже складно візуально розрізнити амплітудні та часові параметри ЕКГ-сигналу, які треба проаналізувати, і такий зашумлений сигнал не придатний до аналізу.

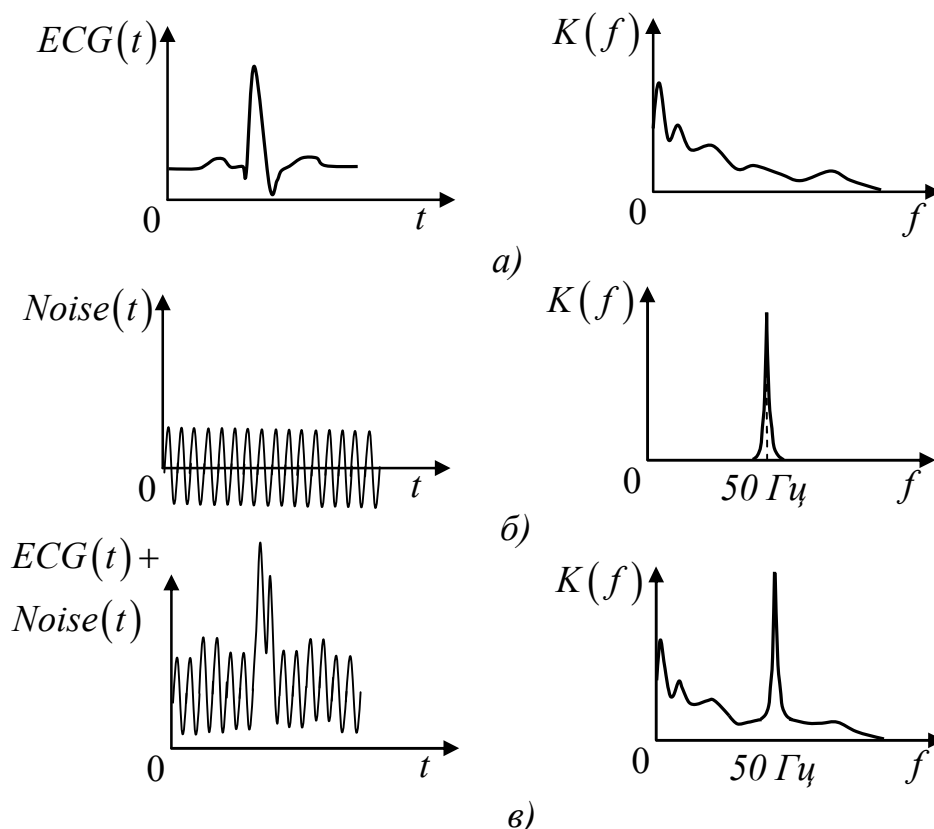


Рисунок 5.1 – а) сигнал та спектр ЕКГ, б) синусоїдальна мережева перешкода 50 Гц та її спектр, в) зашумлений сигнал ЕКГ та його спектр

Для того, щоб позбутися зайвих складових та очистити сигнал від небажаного шуму, необхідно виконувати попередню обробку сигналу. Треба створити систему, яка б при подачі на її вхід зашумленого сигналу, видавала на виході лише ту корисну складову, яка потрібна нам для подальшого аналізу. Така система називається фільтром.

5.2. Типи та характеристики фільтрів

Опис процесу фільтрації і роботу фільтрів зручно розглядати в частотній області із залученням понять спектрального розкладу за Фур'є та частотних характеристик систем. В цьому посібнику ми обмежимося лінійною неадаптивною фільтрацією, тобто всі фільтри будуть лінійними стаціонарними системами. Основна ідея та передумова лінійної неадаптивної фільтрації в тому, що в спектрі зашумленого сигналу існують частотні складові, які відповідають корисній складовій, а також такі, що відповідають шумовій складовій. І області частот, на яких присутні ці складові, не перекриваються.

Фільтрація – це обробка сигналу в часовій області, в результаті якої змінюється спектральний склад початкового сигналу. Зміни зазвичай полягають в зменшенні або вилученні певних складових в спектрі зашумленого сигналу. Вилучатися повинні

лише ті спектральні складові, які вважаються такими, що відносяться до шуму в зашуменому аналізованому сигналі. Водночас, інші спектральні складові, які відносяться до корисного сигналу, повинні залишатися недеформованими при проходженні через фільтр.

Фільтр – це система, яка виконує частотно-залежну обробку сигналів, змінюючи форму сигналу шляхом зменшення амплітуд деяких його спектральних складових.

Метою фільтрації є покращення якості сигналу (вилучення або зменшення шумової складової), виділення з сигналів інформації або частотного розділення кількох сигналів, які раніше були поєднані для ефективного використання доступного каналу зв'язку. Для того, щоб виконати фільтрацію якісно та досягти мети – позбутися від шуму – треба заздалегідь знати спектральні характеристики як корисного сигналу, так і шумової складової.

Розглянемо основні відомості щодо фільтрації неперервних сигналів. Фільтр є, як правило, лінійною системою, яка може бути описана своєю імпульсною характеристикою. Якщо на вхід фільтра з імпульсною характеристикою $h(t)$ подати сигнал $x(t)$, то вихідний сигнал цього фільтра можна отримати, розрахувавши згортку:

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau) h(t - \tau) d\tau.$$

Еквівалентом згортки сигналу і імпульсної характеристики в частотній області є множення їх спектрів, а отже, знаючи спектр вхідного сигналу, можна записати:

$$Y(f) = H(f) X(f),$$

де

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-j2\pi ft} dt$$

$H(f)$ – комплексна частотна характеристика фільтра (frequency response), яка є перетворенням Фур'є його імпульсної характеристики.

Отже, як видно з рівняння, спектр вихідного сигналу фільтра, який ми отримаємо після фільтрації, є результатом множення спектру вхідного сигналу на комплексну частотну характеристику (КЧХ) фільтра. В залежності від того, які властивості має КЧХ фільтра (для яких частот вона має великі значення коефіцієнту передачі, а для яких – малі), так система і змінить спектр вхідного сигналу.

КЧХ – це комплексна функція, яка може бути представлена в алгебраїчній формі:

$$H(f) = H_R(f) + jH_I(f)$$

та в експоненціальній формі:

$$H(f) = |H(f)| e^{j\angle H(f)} = K(f) e^{j\varphi(f)},$$

де $K(f)$ – амплітудно-частотна характеристика (АЧХ, amplitude response);

$\varphi(f)$ – фазо-частотна характеристика (ФЧХ, phase response).

За типами сигналів, які обробляє фільтр, фільтри бувають аналогові та цифрові.

Аналоговий фільтр – електрична схема, що реалізує потрібну АЧХ та ФЧХ фільтра, на вхід якої подається неперервний сигнал, що треба відфільтрувати.

Цифровий фільтр – це математичний алгоритм фільтрації, реалізований на апаратному або програмному рівні, який на вході має дискретний вхідний сигнал та генерує дискретний вихідний сигнал. Цифровий фільтр – це дискретна система, яка має потрібну АЧХ та ФЧХ, що відповідає одному з типів фільтрів (рис. 5.2). Особливістю цифрового фільтра є те, що сигнал подається на вхід в оцифрованому вигляді.

За формою графіку модуля комплексної частотної характеристики (за видом АЧХ) фільтри ділять на чотири основні типи:

- фільтр нижніх частот (ФНЧ, low-pass filter) – має порівняно великий коефіцієнт передачі на частотах, які знаходяться в околі нульової частоти, і порівняно низький – на інших частотах. Тому такі фільтри зберігають незмінною величину спектральних складових вхідного сигналу на нижніх частотах; говорять, що ФНЧ пропускають низькочастотні сигнали та не пропускають високочастотні;

- фільтр верхніх частот (ФВЧ, high-pass filter) – має високий коефіцієнт передачі на частотах, які більші за деяку частоту f_p , і низький – на частотах від нуля до f_p . Отже, вони пропускають без змін частину спектру вхідного сигналу, яка знаходиться на частотах від f_p до $+\infty$ і не пропускають складові спектру сигналу від 0 до f_p . ФВЧ пропускають високочастотні сигнали;

- смуговий фільтр (СФ, band-pass filter) – має великий коефіцієнт передачі лише в певній смузі частот між частотами f_{p1} та f_{p2} . Поза цим проміжком коефіцієнт передачі малий. Отже, на вихід такого фільтра пройдуть лише ті спектральні складові сигналу, які знаходяться в смузі пропускання фільтра, і не пройдуть ті, які лежать від нуля до f_{p1} та від f_{p2} до $+\infty$;

- загороджувальний фільтр (ЗФ, band-stop filter) – є дуальним до смугового: великий коефіцієнт передачі в нього поза проміжком частот від f_{p1} до f_{p2} . Починаючи від нульової частоти і до f_{p1} , а також від f_{p2} до $+\infty$ коефіцієнт передачі великий. Тому при проходженні через такий фільтр сигналу, із його спектру видаляться складові, що лежать між f_{p1} та f_{p2} , а всі інші складові пройдуть на вихід фільтру без змін. Іноді загороджувальний фільтр називають режекторним фільтром. Якщо смуга затримки такого фільтра порівняно вузька, такий загороджувальний фільтр називають іноді фільтром-пробкою (notch-filter). Зокрема, такі фільтри використовують для позбавлення від мережевої перешкоди, коли треба вирізати зі спектру сигналу лише спектральну складову в вузькому околі частоти 50 Гц.

На рисунку 5.2 наведені АЧХ «ідеальних» фільтрів чотирьох типів.

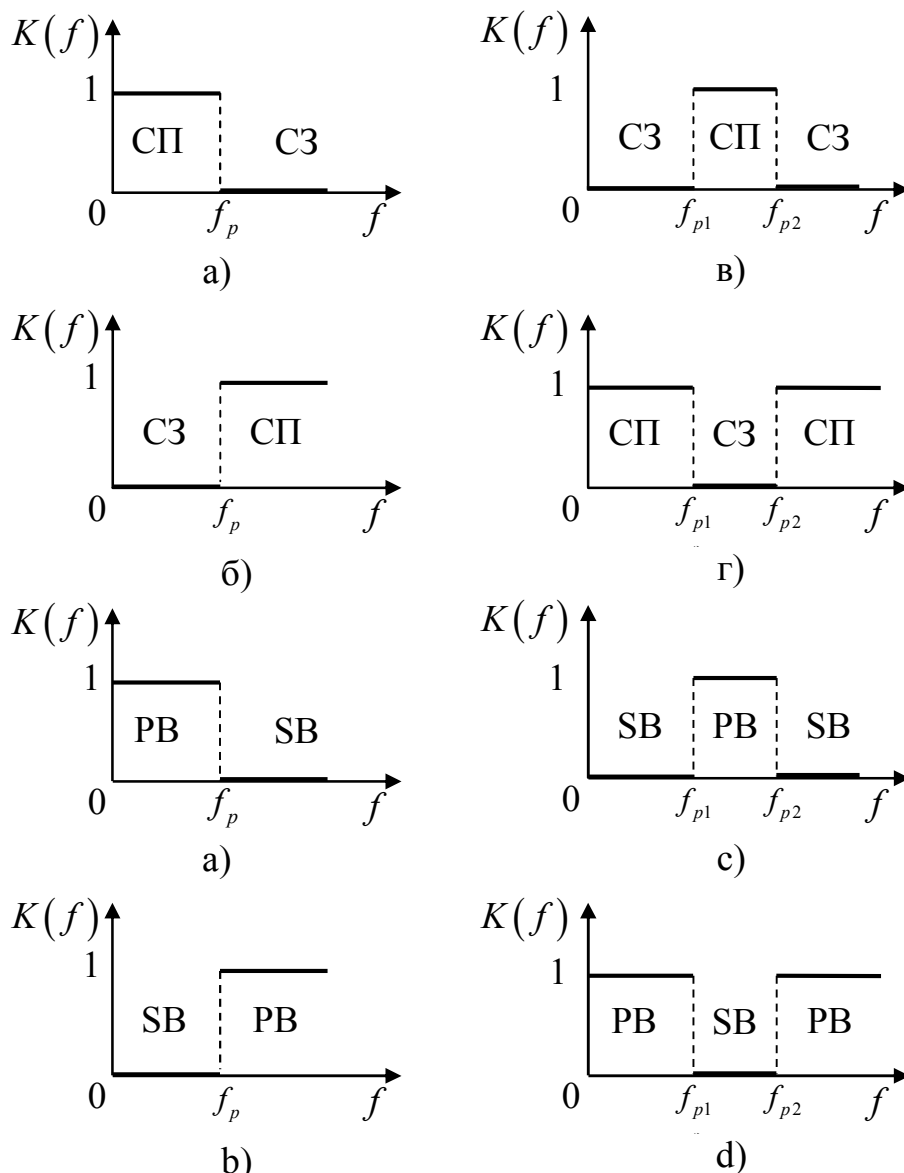


Рисунок 5.2 – АЧХ ідеальних фільтрів: а) ФНЧ, б) ФВЧ, в) СФ, г) ЗФ

Проміжок частот, для яких коефіцієнт передачі фільтра є великим, називається *смугою пропускання* (СП, passband). Якщо спектральні складові вхідного сигналу фільтра потрапляють в смугу пропускання, то вони не зазнають ослаблення, проходять на вихід фільтра без змін та залишаються в спектрі вихідного сигналу.

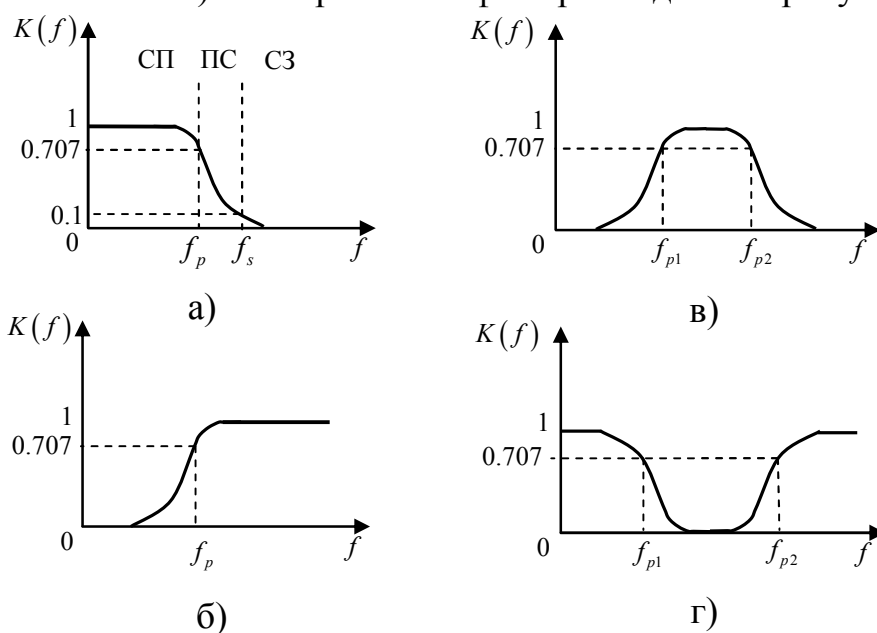
Проміжок частот, для яких коефіцієнт передачі фільтра малий (для ідеального фільтра він дорівнює нулю), називається *смугою затримки* (СЗ, stopband). Якщо частина спектру сигналу потрапляє в смугу затримки, то ці складові зазнають значного ослаблення, і у вихідному сигналі фільтра вони присутні не будуть. В цьому випадку говорять, що ця частина спектру сигналу затримана фільтром.

Частота зрізу f_p для ідеальних фільтрів – це частота, яка розмежовує смуги пропускання і затримки фільтра. Для ФНЧ та ФВЧ вона одна, а СФ та ЗФ мають дві частоти зрізу.

Значення коефіцієнту передачі фільтра в смузі пропускання, як правило, намагаються задати рівним 1. Це значить, що амплітуда тих частотних складових вхідного сигналу, які потрапляють в смугу пропускання, помножується на 1, тобто не змінюється при проходженні з входу на вихід фільтра. Таким способом частина спектру сигналу в смузі пропускання не спотворюється при фільтрації. Якщо фільтр спроектовано правильно, щоб у смузі затримки потрапляли лише частотні складові шуму, то в результаті проходження сигналу через фільтр шумова складова сигналу зникне, а залишиться лише корисна складова.

Фільтри з рисунку 5.2 називаються ідеальними, оскільки на частоті зрізу відбувається нескінченно швидка зміна коефіцієнта передачі фільтра від нуля до 1 (фільтр не пропускає частотну складову зовсім) до 1 (фільтр пропускає частотну складову без спотворень). Реалізувати такі фільтри у вигляді реальних систем неможливо, оскільки лінійна система з такою АЧХ є недетермінованою (імпульсна характеристика такого фільтра повинна починатися на $-\infty$).

В реальних лінійних системах-фільтрах неможливо забезпечити різкий перехід між смугою пропускання та смугою затримки. Між цими двома смугами буде знаходитися частотний проміжок, на якому коефіцієнт передачі змінюється – перехідна смуга (ПС, transition band). АЧХ реальних фільтрів подано на рисунку 5.3.



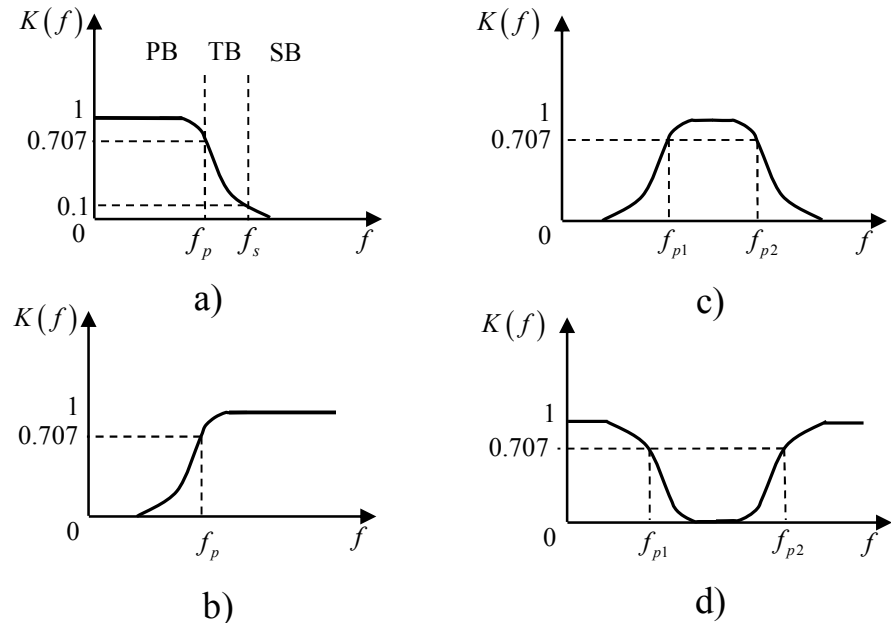


Рисунок 5.3 – АЧХ реальних фільтрів: а) ФНЧ, б) ФВЧ, в) СФ, г) ЗФ

Видно, що зміна коефіцієнта передачі для таких фільтрів відбувається не раптово. Смуга пропускання реальних фільтрів визначається по-іншому: це смуга, в якій фільтр має достатньо великий коефіцієнт передачі. Вважається, що це смуга частот, для яких коефіцієнт передачі зменшується не менше, ніж в $\sqrt{2}$ разів.

На основі цього визначається *частота зрізу* фільтра – це та частота, на якій коефіцієнт передачі зменшується в $\sqrt{2}$ разів відносно максимального коефіцієнта передачі в смузі пропускання. Для випадку, коли коефіцієнт передачі фільтра вимірюється в децибелах, то частота зрізу визначається на рівні -3 дБ від максимального значення.

Для реальних фільтрів також визначають ще одну характеристику – частоту затримки f_s . Це частота, яка відокремлює смугу затримки від перехідної смуги. Для задавання цієї частоти немає усталеного підходу (аналогічного до вибору частоти зрізу на рівні -3 дБ), тому її можна обирати більш-менш довільно виходячи із задачі фільтрації. В цьому випадку дослідник самостійно обирає, для яких значень коефіцієнтів передачі фільтра можна вважати, що спектральні складові сигналу ослаблюються достатньо сильно. Іноді вважають, що смуга затримки починається там, де коефіцієнт передачі фільтра зменшується в 10 разів відносно максимального.

5.2. Етапи розробки фільтрів

Розробка фільтра виконується в кілька етапів.

1. Специфікація вимог до фільтра – визначення бажаних властивостей того фільтра, який би вирішив поставлену задачу фільтрації. Крім того, на етапі специфікації, як правило, задаються такі вимоги до фільтра:

а) тип джерела та приймача сигналів, інтерфейс вводу-виводу, швидкість передачі даних, найбільша частота в спектрі сигналу, що становить інтерес;

б) характеристика фільтра (бажана АЧХ, ФЧХ). Вимоги до АЧХ формуються в частотній області (частоти зрізу та затримки, допустимі пульсації коефіцієнта передачі та т.і.), також задаються максимальні припустимі відхилення від вимог;

в) принцип реалізації: комп'ютерна програма, система цифрової обробки сигналів на базі процесора, аналогова схема. Визначаються також, в реальному або у відкладеному часі повинна виконуватися фільтрація;

г) інші вимоги до структури, наприклад, ціна комплектуючих, вага готового пристрою або блоку, габаритний розмір, стійкість до температури або радіації та т.і.

2. Апроксимація специфікацій – вибір частотних характеристик реальної системи, яка буде реалізовуватися. Визначення коефіцієнтів фільтра, квантування коефіцієнтів.

3. Представлення фільтра належною структурою.

4. Моделювання фільтра – перевірка, чи задовольняє отриманий фільтр заданим вимогам із застосуванням програмних пакетів.

5. Реалізація фільтра на апаратному або програмному рівні – розробка схеми або програми, яка реалізує потрібні частотні характеристики.

6. Оптимізація отриманого фільтра – аналіз роботи фільтра, виявлення недоліків та покращення фільтра шляхом підбору оптимальної схеми та параметрів фільтра.

5.3. Поняття про апроксимацію фільтрів

Як правило, вимоги до фільтра задаються у вигляді параметрів ідеального фільтра, АЧХ якого має вигляд, наведений на рисунку 5.2. Як було сказано раніше, створити реальну систему з такими АЧХ неможливо, отже, розроблений реальний фільтр буде деяким наближенням до бажаного ідеального.

На етапі апроксимації спочатку задається вигляд АЧХ ідеального фільтра (тип фільтра та частота зрізу). Потім підбирається аналітичний опис функції в частотній області, яка апроксимує ідеальну АЧХ виходячи з певних критеріїв щодо похибки апроксимації. Після того, як апроксимуюча функція отримана в операторній формі

(через оператор Лапласа), її записують у вигляді раціонального дробу, в чисельнику і в знаменнику якого є поліноми. На цьому етапі буде отримано математичний опис передавальної функції реального аналогового фільтра, АЧХ якого апроксимує АЧХ бажаного ідеального фільтра.

З теорії електронних кіл відомо, що передавальна функція лінійної системи може бути записана в операторній формі через оператор Лапласа

$$H(p) = \frac{Y(p)}{X(p)} = \frac{B_0 + B_1p + B_2p^2 + \dots + B_np^n}{A_0 + A_1p + A_2p^2 + \dots + A_mp^m} = \frac{(b_{01} + b_{11}p)(b_{02} + b_{12}p + a_{22}p^2) \dots (b_{0n} + b_{1n}p + b_{2n}p^2)}{(a_{01} + a_{11}p)(a_{02} + a_{12}p + a_{22}p^2) \dots (a_{0m} + a_{1m}p + a_{2m}p^2)} \quad (5.1)$$

у вигляді дробу, чисельник та знаменник якого є добутками поліномів не вище другого ступеню. Сукупність значень нулів чисельника називають нулями передавальної функції, сукупність значень нулів знаменника – полюсами. *Порядок фільтра* – це кількість полюсів передавальної функції фільтра.

Задача апроксимації полягає у знаходженні порядку фільтра, а також всіх коефіцієнтів передавальної функції фільтра.

Для того, щоб створити реальний аналоговий фільтр, вираз (5.1) доцільно представити у вигляді добутків множників не вище другого порядку:

$$H(p) = k_1(p) \cdot k_2(p) \cdot \dots \cdot k_n(p)$$

Кожний з множників є передавальною функцією деякої простої схемної ланки. Тобто, аналоговий фільтр будь-якого порядку можна отримати при послідовному з'єднанні фільтрів першого та другого порядку (аналогова схемотехніка).

Для того, щоб описати математично АЧХ фільтрів, здебільшого використовують поліноміальну апроксимацію. Види апроксимації часто названі по імені тих дослідників, що розробили або запропонували поліноми, які використовуються для апроксимації.

Фільтри з апроксимацією за Баттервортом

Особливістю апроксимації за Баттервортом є те, що АЧХ фільтра над смугою пропускання є максимально плоскою (рис. 5.4). Це значить, що вона не буде спотворювати низькочастотну частину спектру сигналу, який пройде через фільтр. Інша властивість отриманої апроксимації – монотонність АЧХ як в смузі пропускання, так і в смузі затримки. Квадрат АЧХ ФНЧ з апроксимацією за Баттервортом має вигляд:

$$|H(\omega)|^2 = \frac{1}{1 + \left(\frac{\omega}{\Omega_c}\right)^{2n}},$$

де Ω_c – частота зрізу;

n – порядок фільтра.

Недоліками апроксимації за Баттервортом є низька вибірковість в частотній області через те, що перехідна смуга цього фільтра є широкою, а крутизна характеристики в перехідній смузі є низькою. Чим більший порядок фільтра, тим крутизна спаду АЧХ в перехідній смузі більша. Отже, для підвищення вибірковості часто необхідно брати велику кількість ланок при схемній реалізації фільтра, що не завжди зручно.

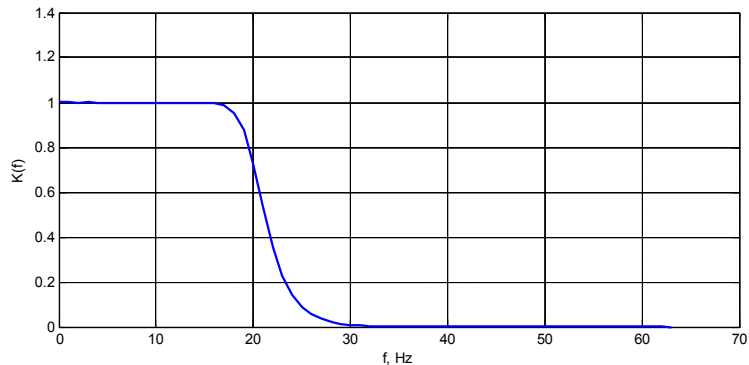


Рисунок 5.4 – АЧХ ФНЧ з апроксимацією за Баттервортом

Фільтри з апроксимацією за Чебишовим I і II роду

АЧХ фільтрів Чебишова I роду рівномірно пульсує над смугою пропускання та рівномірна над смугою затримки, а II роду – над смугою пропускання є рівномірною, а в смузі затримки має пульсації (рис. 5.5). Квадрат АЧХ фільтру Чебишова I роду має вигляд:

$$|H(\omega)|^2 = \frac{1}{1 + \varepsilon^2 V_n^2\left(\frac{\omega}{\Omega_c}\right)},$$

де $V_n(x)$ – поліном Чебишова n -го порядку:

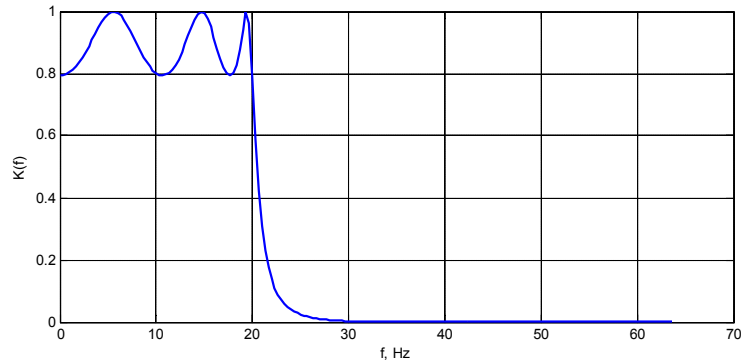
$$V_n(x) = \begin{cases} \cos(n \cdot \arccos(x)), & |x| \leq 1 \\ ch(n \operatorname{Arch}(x)), & |x| > 1 \end{cases},$$

Функція коливається між 1 та $\frac{1}{1 + \varepsilon^2}$ в межах від 0 до Ω_c та монотонно спадає при $\omega > \Omega_c$.

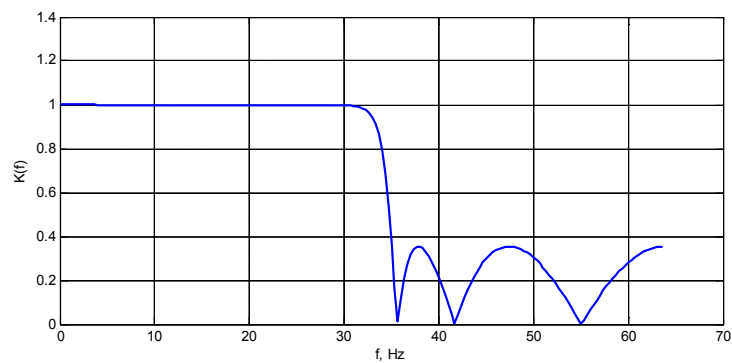
Для побудови фільтра потрібно задати три параметри: ε , Ω_c , n . В реальних задачах ε визначається допустимими пульсаціями в смузі пропускання. Після задавання цих параметрів підбирають прийнятний порядок фільтра виходячи з необхідної смуги затримки.

Квадрат АЧХ фільтру Чебишова II роду має вигляд:

$$|H(\omega)|^2 = \frac{1}{1 + \varepsilon^2 \left[\frac{V_n(\Omega_c)}{V_n\left(\frac{\Omega_c}{\omega}\right)} \right]^2}. \quad (5.2)$$



а)



б)

Рисунок 5.5 – АЧХ ФНЧ з апроксимацією за Чебишовим I роду (а) та II роду (б)

Еліптичні фільтри (фільтри з апроксимацією Кауера)

Для такої апроксимації пульсації будуть присутні і в смузі пропускання, і в смузі затримки (рис. 5.6). Можна довести, що такий тип апроксимації є оптимальним в сенсі, що при фіксованому порядку фільтра та значеннях частоти затримки і допустимих коливань в смугах затримки та пропускання, ширина перехідної смуги виявляється найменшою серед усіх можливих апроксимацій АЧХ.

Квадрат АЧХ має вигляд:

$$|H(\omega)|^2 = \frac{1}{1 + \varepsilon^2 U_n^2(\omega, R)}, \quad (5.3)$$

де $U_n(\omega)$ – раціональна функція Чебишова (еліптична функція Якобі);

R – параметр, який характеризує пульсації.

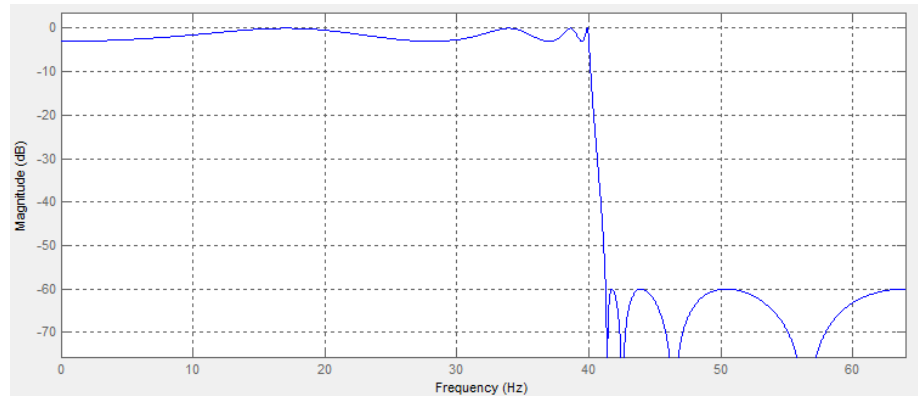


Рисунок 5.6 – АЧХ ФНЧ з апроксимацією за Кауером

Варто зауважити, що іноді процес отримання наближеного опису АЧХ називають скорочено «апроксимацію фільтрів», але правильна назва – апроксимація передавальної функції фільтра. Відбувається наближення бажаної (ідеальної) передавальної функції реальною, яку можна і треба буде реалізувати на практиці. Так само, іноді говорять скорочено про «фільтр Баттерворта», «фільтр Чебишова» та т.і., хоча повністю вірним є говорити про «фільтр, передавальна функція якого отримана в результаті апроксимації АЧХ поліномом Баттерворта».

Також треба відмітити, що фільтри всіх типів проектуються з використанням низькочастотного прототипу – деякого еквівалентного ФНЧ фільтра для того фільтра, який потрібно спроектувати. Від специфікації необхідного фільтра переходять до специфікації НЧ-прототипу, апроксимують його АЧХ та за допомогою процедури перетворення частот отримують коефіцієнти передавальної функції потрібного фільтра. Якщо необхідно розробити аналоговий фільтр, то на цьому етапі починається його реалізація.

Докладніше проектування аналогових фільтрів вивчає дисципліна «Аналогова схемотехніка». Якщо потрібний цифровий фільтр, то після того, як отримано аналоговий прототип (тобто опис передавальної функції аналогового фільтра з потрібними властивостями), застосовуються різні методи проектування цифрових фільтрів за аналоговим прототипом. Про ці методи буде іти мова далі.

5.4. Цифрові БІХ- та КІХ- фільтри

Цифровий фільтр призначений для фільтрації дискретних сигналів. Як правило, це лінійна дискретна система, яка, незалежно від реалізації (у вигляді комп'ютерної програми чи електронного пристрою) може бути описана одним з чотирьох способів: різницеvim рівнянням, імпульсною характеристикою, характеристичною функцією або комплексною частотною характеристикою.

Цифрові фільтри за властивостями імпульсної характеристики $h[n]$ ділять на два класи: з кінцевою імпульсною характеристикою (*КІХ-фільтри*) та з безкінечною імпульсною характеристикою (*БІХ-фільтри*).

Вихідний сигнал КІХ-фільтра можна отримати за такими рівняннями:

$$y[n] = \sum_{k=0}^{N-1} h[k]x[n-k], \quad (5.4)$$

$$H(z) = \sum_{k=0}^{N-1} h(z)z^{-k},$$

де $h[n]$ – відліки імпульсної характеристики;

N – довжина фільтра, тобто кількість коефіцієнтів фільтра.

Вираз (5.4) – різницеве рівняння КІХ-фільтра, воно представляє роботу фільтра в часовій області в його нерекурсивній формі. Кількість доданків в сумі є скінченною, тому для стійкого фільтра можливо розрахувати вихідний сигнал, маючи імпульсну характеристику.

Для БІХ-фільтра вихідний сигнал формально отримується також через рівняння згортки:

$$y[n] = \sum_{k=0}^{\infty} h[k]x[n-k],$$

але на практиці розрахунок за цим рівнянням неможливий, оскільки тривалість імпульсної характеристики нескінченна. Для практичних застосувань різницеве рівняння для БІХ-фільтра треба записати в рекурсивній формі:

$$y[n] = \sum_{k=0}^{\infty} h[k]x[n-k] = \sum_{k=0}^{M-1} b_k x[n-k] - \sum_{k=1}^{N-1} a_k y[n-k].$$

При такому записі різничевого рівняння кількість доданків в обох сумах скінченна, а, отже, можна порахувати значення вихідного сигналу БІХ-фільтра.

Передавальна функція БІХ-фільтра може бути записана таким чином:

$$H(z) = \frac{b_0 + b_1 z + b_2 z^{-1} + \dots + b_N z^{-N}}{1 + a_1 z + a_2 z^{-1} + \dots + a_N z^{-M}} = \frac{\sum_{k=0}^{M-1} b_k z^{-k}}{1 + \sum_{k=1}^{N-1} a_k z^{-k}}.$$

5.5. Основні відомості щодо методів розрахунку коефіцієнтів цифрових фільтрів

Для того, щоб виконувати фільтрацію дискретних сигналів, необхідно мати коефіцієнти фільтра (КІХ- та БІХ фільтри), або відліки імпульсної характеристики фільтра (для КІХ-фільтрів).

Розрахунок КІХ-фільтрів

Існують три основні методи: метод зважування, оптимальний метод (ґрунтується на мінімізації похибки відхилення реальної АЧХ від ідеальної, використовується алгоритм Ремеза) та метод частотної вибірки (допускає рекурсивну реалізацію КІХ-фільтрів, аналогічно до метода зважування, але дискретизація виконується в частотній області і потім робиться перетворення Фур'є).

Метод зважування.

ґрунтується на тому, що імпульсна характеристика та КЧХ фільтра пов'язані оберненим перетворенням Фур'є. Спочатку в часовій області отримують імпульсну характеристику потрібного аналогового фільтра-прототипа. Вона є неперервною функцією часу, в загальному випадку – нескінченною. Для того, щоб отримати скінченну імпульсну характеристику, ідеальна імпульсна характеристика помножається на деяку віконну (вагову) функцію скінченної тривалості, при цьому отримана імпульсна характеристика буде гладко затухати до нуля.

Для отримання відліків імпульсної характеристики дискретного фільтра проводиться дискретизація неперервної імпульсної характеристики аналогового фільтра. Те, наскільки властивості отриманого дискретного фільтра будуть відповідати властивостям початкового фільтра-прототипа, буде залежати від того, наскільки точно *форма* отриманої дискретної імпульсної характеристики буде відображати форму імпульсної характеристики початкового аналогового фільтра-прототипа. Це буде залежати від двох факторів:

- наскільки широкою є ділянка імпульсної характеристики аналогового фільтра, яка вирізана віконною функцією;
- наскільки багато відліків імпульсної характеристики неперервного фільтра буде взято для формування імпульсної характеристики дискретного фільтра.

КІХ-фільтр може бути реалізований як у вигляді рекурсивної, так і нерекурсивної системи.

Розрахунок БІХ-фільтрів

БІХ-фільтр може бути реалізований тільки як система зі зворотним зв'язком, тобто рекурсивна система. Існують кілька методів проектування БІХ-фільтрів:

- *метод розміщення нулів та полюсів* – полягає в безпосередньому задаванні розташування полюсів та нулів на комплексній площині для отримання потрібної КЧХ. Після того, як нулі та полюси розміщені, записується вираз КЧХ цифрового фільтра;
- *метод інваріантного перетворення імпульсної характеристики* – за допомогою перетворення Лапласа з підходящої аналогової передавальної функції $H(s)$ отримують імпульсну характеристику аналогового фільтра $h(t)$, після цього її дискретизують. Отримана дискретна функція $h(nT_s)$ піддається

z -перетворенню і дає в результаті бажану характеристичну функцію цифрового фільтра $H(z)$, з якої отримують КЧХ та різницеве рівняння цифрового фільтра;

– *узгоджене z -перетворення* – при цьому кожен полюс та нуль на безпосередньо переводяться з p -площини (p – оператор Лапласа) на z -площину за правилом: $(p - a) \rightarrow (1 - z^{-1}e^{aT_s})$;

– *білінійне z -перетворення*.

Білінійне z -перетворення (метод дробово-лінійного перетворення).

Це один з найважливіших методів побудови БІХ-фільтрів. Цей метод ґрунтується на перетворенні змінних p (оператора Лапласа) та z , при якому вся пряма $j\omega$ в p -площині один раз огортається навколо одиничної окружності в z -площині. Закон перетворення нелінійний, білінійне перетворення задається формулою:

$$p = \frac{2}{T_s} \left(\frac{1 - z^{-1}}{1 + z^{-1}} \right), \text{ або } p = \frac{1 - z^{-1}}{1 + z^{-1}}.$$

Цей вираз підставляється в передавальну функцію аналогового фільтра в операторній формі, і можна отримати характеристичну функцію цифрового фільтра в z -області.

Недоліком цього методу є нелінійність перетворення частотних вісей, що проявляється у спотвореннях АЧХ та ФЧХ фільтра. Отже, перетворення можна застосовувати лише в разі, коли такі спотворення є припустимими або їх можна компенсувати. Як правило, такий метод використовують для проектування ФВЧ, ФНЧ та СФ з кусково-постійною АЧХ.

5.5. Побудова структур цифрових фільтрів

Робота фільтра в часовій області описується різницеvim рівнянням, яке показує, як отримати поточний відлік вихідного сигналу фільтра, маючи відліки вихідного та вхідного сигналу:

$$\sum_{k=0}^{N-1} a_k y[n-k] = \sum_{m=0}^{M-1} b_m x[n-m],$$

$$y[n] = \sum_{m=0}^{M-1} b_m x[n-m] - \sum_{k=1}^{N-1} a_k y[n-k].$$

Іноді для того, щоб більш наочно представити, як саме працює цифровий фільтр, його різницеве рівняння представляється візуально у вигляді структури.

Структура фільтра відображає види та послідовність математичних операцій, які виконуються над відліками сигналів.

Як видно з різницевого рівняння, для того, щоб розрахувати значення вихідного сигналу $y[n]$, необхідно виконувати такі операції:

– *додавання*. Блок в структурній схемі, який виконує додавання, називається суматор; він представлений на рис. 5.7.

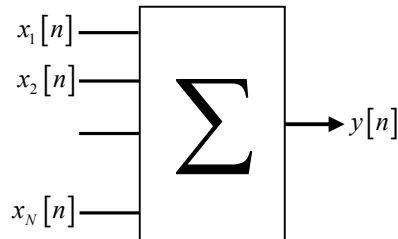


Рисунок 5.7 – Схематичне зображення суматора

Робота суматора описується таким рівнянням:

$$y[n] = x_1[n] + x_2[n] + \dots + x_N[n] = \sum_k x_k[n]$$

– *множення на скаляр*. Блок-перемножувач зображено на рисунку 5.8.

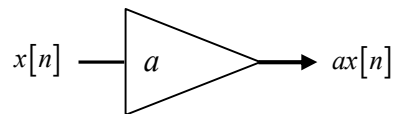


Рисунок 5.8 – Схематичне зображення перемножувача на скаляр

– *затримка на один відлік*. Елемент затримки зображено на рисунку 5.9.

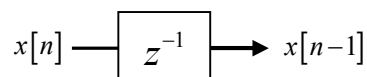


Рисунок 5.9 – Схематичне зображення елемента затримки на один відлік

Побудова структури фільтра за заданим різницевою рівнянням зводиться до відображення кожної дії над відліками вхідного та вихідного сигналів за допомогою відповідних блоків.

Наприклад, побудуємо структуру нерекурсивного фільтра, який описується різницевою рівнянням:

$$y[n] = 2x[n] + 3x[n-1] - 5x[n-3].$$

Із рівняння видно, що кожний відлік вихідного сигналу $y[n]$ складається з суми трьох доданків: помноженого на 2 поточного відліка вхідного сигналу $2x[n]$,

помноженого на 3 відліка вхідного сигналу, що затриманий на один такт $3x[n-1]$, та помноженого на -5 відліку вхідного сигналу, який затриманий на три такти $-5x[n-3]$. Структура фільтра подана на рисунку 5.10.

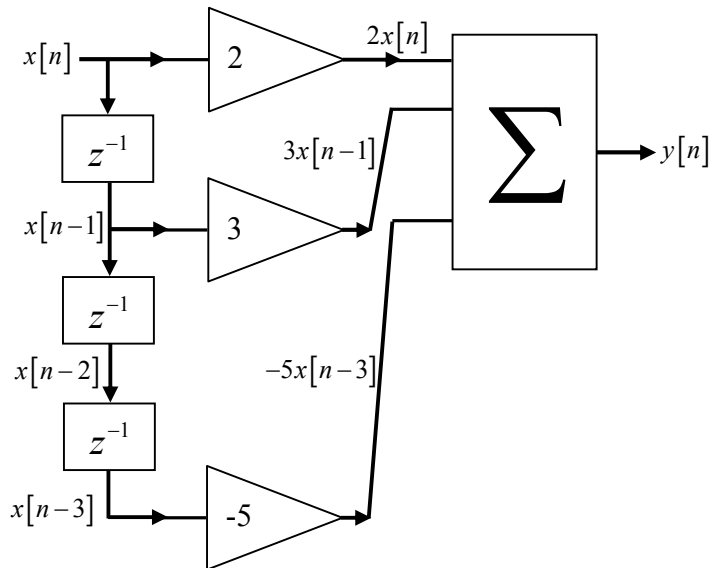


Рисунок 5.10 – Приклад структури нерекурсивного фільтра

У випадку, коли фільтр рекурсивний, на суматор подаються не тільки затримані і помножені на константи відліки вхідного сигналу, а також і затримані відліки вихідного сигналу. На рисунку 5.11 для прикладу подано структуру рекурсивного фільтра з різницею рівнянням:

$$y[n] = 3x[n] + 2x[n-1] - 5y[n-1] + 4y[n-2] + 8y[n-3]. \quad (5.5)$$

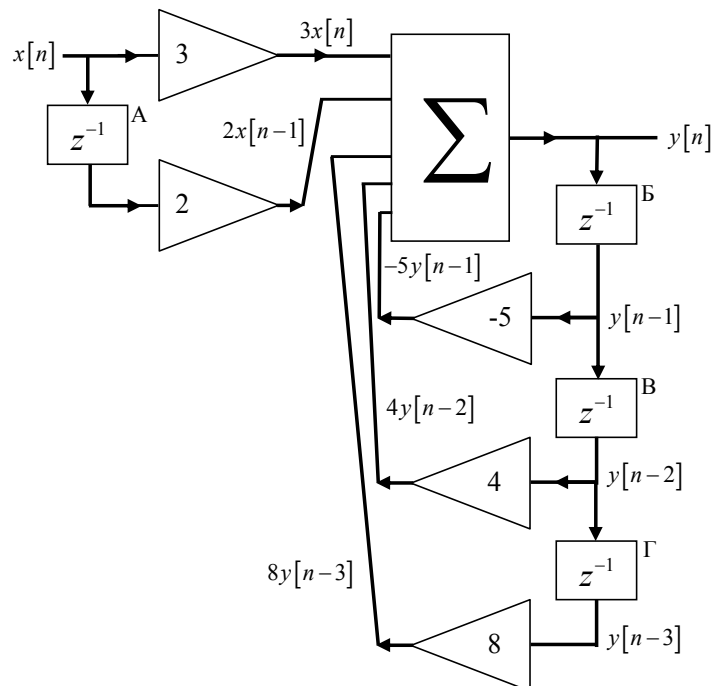


Рисунок 5.11 – Приклад структури рекурсивного фільтра

Аналізуючи структуру фільтра, наочно видно, як відбувається формування вихідного сигналу, і яку затримку дає фільтр. Проаналізуємо, як відбуваються розрахунки значень вихідного сигналу фільтра (5.5).

Нехай маємо нульові початкові умови, тобто у всіх елементах затримки зберігаються нульові значення до того, як на вхід подається вхідний сигнал.

В початковий момент часу $n=0$ на вхід фільтра поступає початковий відлік вхідного сигналу $x[0]$. Він домножається на 3 та проходить на вихід фільтра, формуючи відлік вихідного сигналу: $y[0]=3x[0]$. Одночасно значення відліку $x[0]$ потрапляє в перший елемент затримки (А), в якому до цього не було ніяких значень. Аналогічно, відлік вихідного сигналу також потрапляє в перший елемент затримки рекурсивної частини (Б) і зберігається там. Отже, в перший момент часу вихідний сигнал $y[0]$ розраховується не по всьому різницевому рівнянню (5.5), а вклад дає тільки перший доданок. Через це даний відлік ще не є відліком правильно відфільтрованого сигналу.

В момент часу $n=1$ на вхід фільтра поступає наступний відлік вхідного сигналу $x[1]$. Він проходить через помножувач та суматор і дає свій вклад у відлік вихідного сигналу фільтра: $3x[1]$. На виході перших елементів затримки в рекурсивній та нерекурсивній частинах в цей момент часу вже з'являються затримані на один такт відліки вхідного ($x[0]$, елемент А) та вихідного ($y[0]$, елемент Б) сигналу, які домножаються на свої коефіцієнти та подаються на суматор. Отже, в результаті маємо на виході $y[1]=3x[1]+2x[0]-5y[0]$. В цьому виразі також ще не присутні всі доданки, які є у (5.5), отже, сигнал ще не фільтрується коректно. В цьому ж такті відбувається оновлення значень в елементах затримки в рекурсивній та нерекурсивній частинах: в елемент А подається значення $x[1]$, в елемент Б подається $y[1]$, а в елемент В передається значення $y[0]$.

В наступний момент часу $n=2$ маємо $x[2]$ на вході та одразу ж доданок $3x[2]$ на виході. Вихідні значення елементів затримки дають доданки $2x[1]$ (елемент А), $-5y[1]$ (елемент Б) та $4y[0]$ (елемент В). В цей момент відбувається оновлення значень в елементах затримки: $x[2]$ в А, $y[2]$ в Б, $y[1]$ в В, $y[0]$ в Г. На виході фільтра маємо значення $y[2]=3x[2]+2x[1]-5y[1]+4y[0]$, що все ще не є коректно відфільтрованим сигналом, оскільки в розрахунках не беруть участь всі доданки з рівняння (5.5).

Нарешті, в момент часу $n=3$ на виході останнього елемента затримки Г з'являється значення $y[0]$, яке дає останній доданок в сумі $8y[0]$. Всі інші елементи затримки та суматори також формують потрібні доданки в сумі (5.5), отже маємо на виході значення:

$$y[3]=3x[3]+2x[2]-5y[2]+4y[1]+8y[0].$$

Видно, що лише в момент часу, коли на вхід прийшов третій доданок, і у всі моменти часу після нього, вихідний сигнал буде розраховуватися по різницевому

рівнянню фільтра з участю всіх доданків. Говорять, що фільтр дає *затримку* вихідного сигналу відносно вхідного. Вона пов'язана з тим, що для того, щоб розрахувати значення вихідного сигналу, необхідно, щоб всі елементи затримки фільтра містили значення сигналу. Ця затримка дорівнює максимальній кількості елементів затримки в рекурсивній або нерекурсивній ланках фільтра. В розглянутому прикладі затримка складає три такти.

5.6. Застосування фільтрації для багатошвидкісної обробки сигналів

Розглянемо одну із задач, яка зустрічається при обробці та аналізі сигналів та для вирішення якої обов'язково необхідно використовувати фільтри.

В сучасних цифрових системах доволі розповсюдженою є обробка даних з різними частотами дискретизації. Це необхідно, якщо сигнал дискретизований з одною частотою дискретизації, а цифрова система обробки та аналізу сигналів розрахована на іншу частоту дискретизації (тактову частоту виконання операцій). В цьому випадку необхідно перетворити сигнал, який дискретизований з одною частотою, на сигнал, дискретизований з іншою частотою. По-суті мова іде про *дискретизацію дискретного сигналу*. В теорії сигналів така обробка називається багатошвидкісною обробкою сигналів (multirate processing).

Обробка сигналу на кількох швидкостях дозволяє:

- уникнути використання дорогих аналогових фільтрів захисту від накладення спектрів. Якщо сигнал дискретизується з частотою, набагато більшою за необхідну за теоремою Котельникова, то для обмеження смуги частот з метою уникнення накладення спектрів знадобиться набагато простіший ФНЧ;
- ефективно обробляти в одних і тих самих системах сигнали з різними спектральними характеристиками (більш чи менш широкими спектрами) та з різними частотами дискретизації;
- зменшити необхідну пам'ять для зберігання сигналів та збільшити швидкість передачі сигналів по каналам зв'язку.

Якщо маємо систему, яка працює на одній частоті дискретизації, а сигнал дискретизований з іншою, то найпростішим виходом було б повернутися до аналогового сигналу та дискретизувати його, як потрібно. Але відновлення неперервного сигналу із дискретного є нетривіальною задачею, яку не завжди легко вирішити. Тому такий підхід використовується дуже рідко, натомість вдаються до зміни частоти дискретизації вже дискретизованого сигналу.

Фундаментальними операціями багатошвидкісної обробки сигналів є:

Децимація – зменшення частоти дискретизації (downsampling).

Інтерполяція – збільшення частоти дискретизації (upsampling).

Децимація з цілим кроком

При децимації сигналу $x[n]$ з цілим кроком M отримаємо сигнал $x'[n]$:

$$x'[n] = x[Mn], \quad (5.6)$$

тобто із кожних M відліків початкового сигналу залишається один відлік і відкидається $M-1$ відлік. Це еквівалентно тому, що частота дискретизації початкового сигналу зменшується в M разів:

$$F'_s = \frac{F_s}{M}. \quad (5.7)$$

Розглянемо питання дискретизації початкового сигналу (дискретизованого з частотою F_s) з новою частотою дискретизації F'_s , з огляду на можливість накладання спектрів.

На рисунку 5.12а подано спектр початкового сигналу $x[n]$, який дискретизовано з частотою F_s . Як відомо із теорії дискретизації, цей спектр є періодичним з періодом повторення F_s . У випадку, коли сигнал з таким спектром буде піддано дискретизації ще раз, спектр результуючого сигналу також буде періодичним, і період його повторення буде рівним новій частоті дискретизації F'_s . На рисунку 5.12б фігурною дужкою показано, яка частина спектру такого передискретизованого сигналу буде накладатися, це частина від $\frac{F'_s}{2}$ до F'_s . Для того, щоб спотворення сигналу внаслідок накладання спектрів не відбувалося, необхідно до того, як дискретизувати сигнал з частотою F'_s , виконати фільтрацію. Фільтр повинен бути таким, щоб вилучити частину спектру, яка буде накладатися; для даного сигналу необхідно використати ФНЧ, приклад АЧХ якого наведено на рисунку 5.12в. В результаті фільтрації спектр сигналу буде мати вигляд, який показано на рисунку 5.12г: тут залишилася лише та частина, яка не буде накладатися. Після того, як буде виконана фільтрація сигналу, можна виконувати децимацію з кроком M . В результаті такої дискретизації дискретного (вже відфільтрованого) сигналу, отримаємо сигнал $x'[n]$, спектр якого матиме вигляд, показаний на рисунку 5.12д. В ньому накладання спектрів відсутнє, тому в дискретному сигналі $x'[n]$ коректно відображені властивості початкового дискретного сигналу $x[n]$.

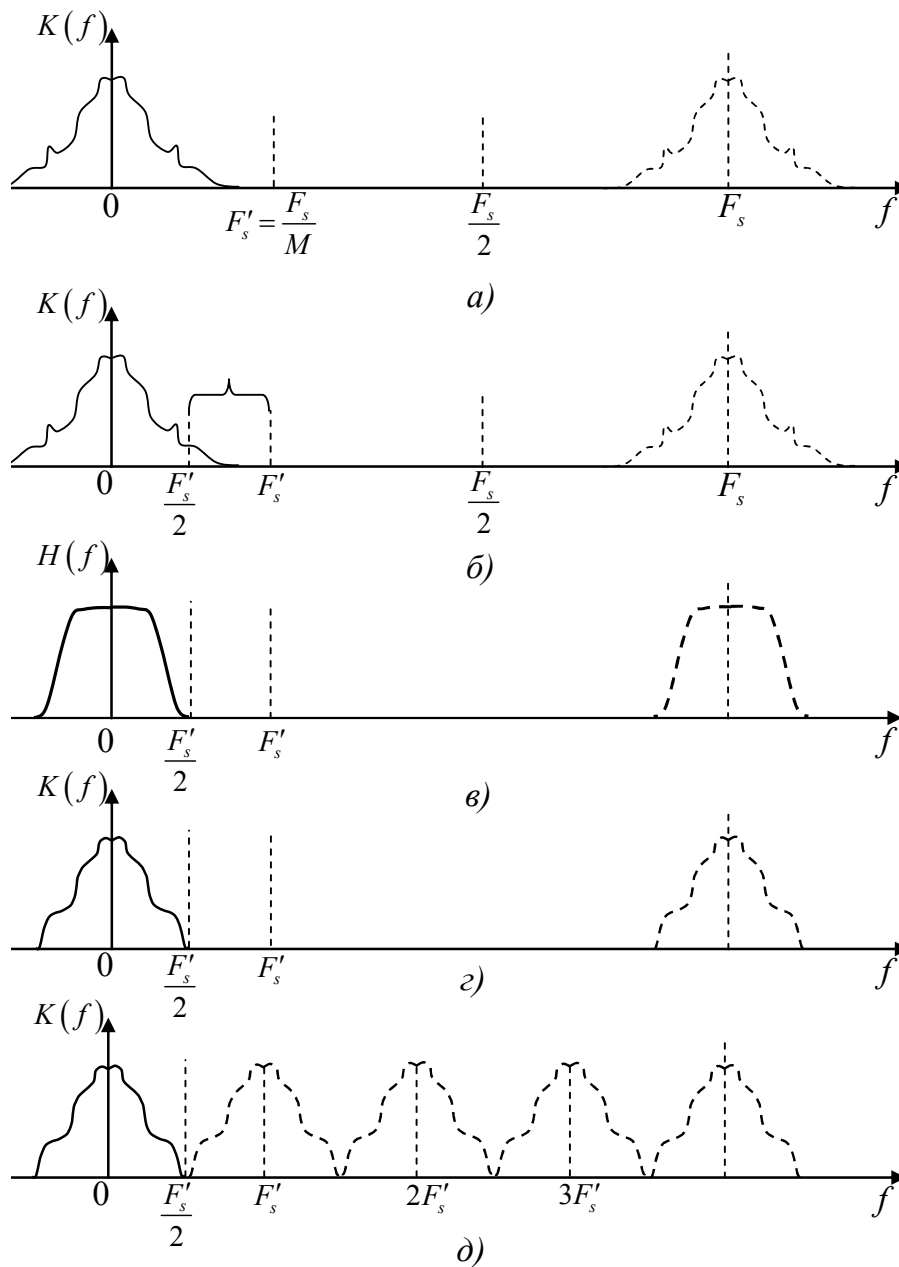


Рисунок 5.12 – а) амплітудний спектр початкового дискретного сигналу $x[n]$, дискретизованого з частотою F_s , б) фігурною дужкою показана частина спектру, що буде накладатися, в) АЧХ фільтра захисту від накладання спектрів, г) спектр початкового сигналу після фільтрації, д) спектр сигналу, дискретизованого з частотою F'_s

Отже, для того, щоб знизити частоту дискретизації дискретного сигналу, треба виконати два кроки: спочатку виконати фільтрацію для запобігання накладенню спектрів, а потім – провести вибірку з кроком M . Блок-схема дециматора подана на рис. 5.13, вона містить ФНЧ та компресор (зменшувач) частоти дискретизації сигналу.

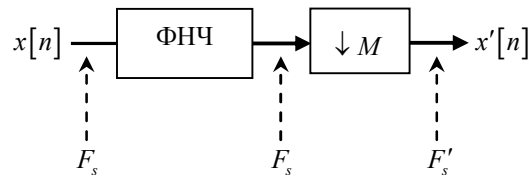


Рисунок 5.13 – Дециматор

В часовій області робота дециматора проілюстрована на рисунку 5.14.

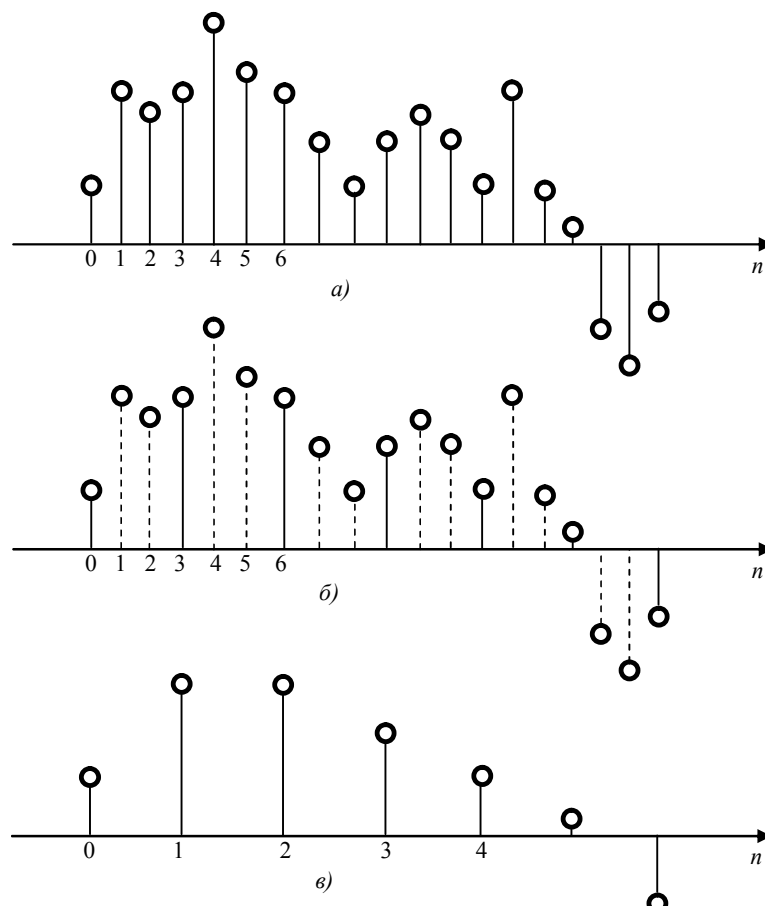


Рисунок 5.14 – Робота дециматора: а) початковий дискретний сигнал $x[n]$ з частотою дискретизації F_s , б) штриховою лінією показані відліки, які будуть пропущені дециматором, в) результуючий дискретний сигнал $x'[n]$ з частотою дискретизації $F'_s = \frac{F_s}{3}$.

Інтерполяція сигналу з цілим кроком

Інтерполяція полягає в тому, що для даного сигналу, який дискретизований з частотою дискретизації F_s , необхідно підвищити частоту дискретизації в ціле число разів:

$$F'_s = LF_s.$$

При цьому на кожний відлік вхідного сигналу треба додати ще $L - 1$ відлік для отримання результуючого сигналу.

Аналогічно до попереднього випадку, необхідно забезпечити відсутність спотворень сигналу при зміні частоти дискретизації внаслідок наявності непотрібних спектральних складових в результуючому сигналі.

На рисунку 5.15а подано спектр сигналу $x[n]$, який дискретизований з частотою F_s . Оскільки сигнал дискретний, то спектр є періодичним, і період повторення дорівнює частоті дискретизації. Форма спектра сигналу $x[n]$ описується його одним періодом, який показаний суцільною лінією, весь спектр сигналу лежить в межах від 0 до половини частоти дискретизації $\frac{F_s}{2}$. На рисунку 5.14б показаний спектр сигналу $x'[n]$ після того, як частота дискретизації була збільшена до $F'_s = LF_s$. З теорії дискретизації відомо, що його спектр буде періодичною функцією частоти, і період буде рівним частоті дискретизації F'_s . Половина одного періоду спектра сигналу тепер буде лежати від 0 до $\frac{F'_s}{2}$. Видно, що на цьому проміжку частот знаходяться не тільки ті спектральні складові початкового сигналу, які можна бачити на 5.14а (від 0 до $\frac{F_s}{2}$), але і зайві складові (показані стрілкою), яких в спектрі початкового сигналу $x[n]$ не було. Вони не описують властивості сигналу, і їх потрібно вилучити при інтерполяції. Для того, щоб це зробити, необхідно залучити фільтрацію даного сигналу. Для цього треба використати фільтр нижніх частот, який пропустить лише ті спектральні складові, які були присутні в початковому сигналі. Приклад АЧХ такого фільтра наведено на рисунку 5.14в. Спектр результуючого сигналу $x'[n]$ після інтерполяції наведено на рисунку 5.14г. Видно, що в ньому присутні лише ті спектральні складові, які були в початковому сигналі $x[n]$, а отже, інтерполяція виконана коректно і без спотворень.

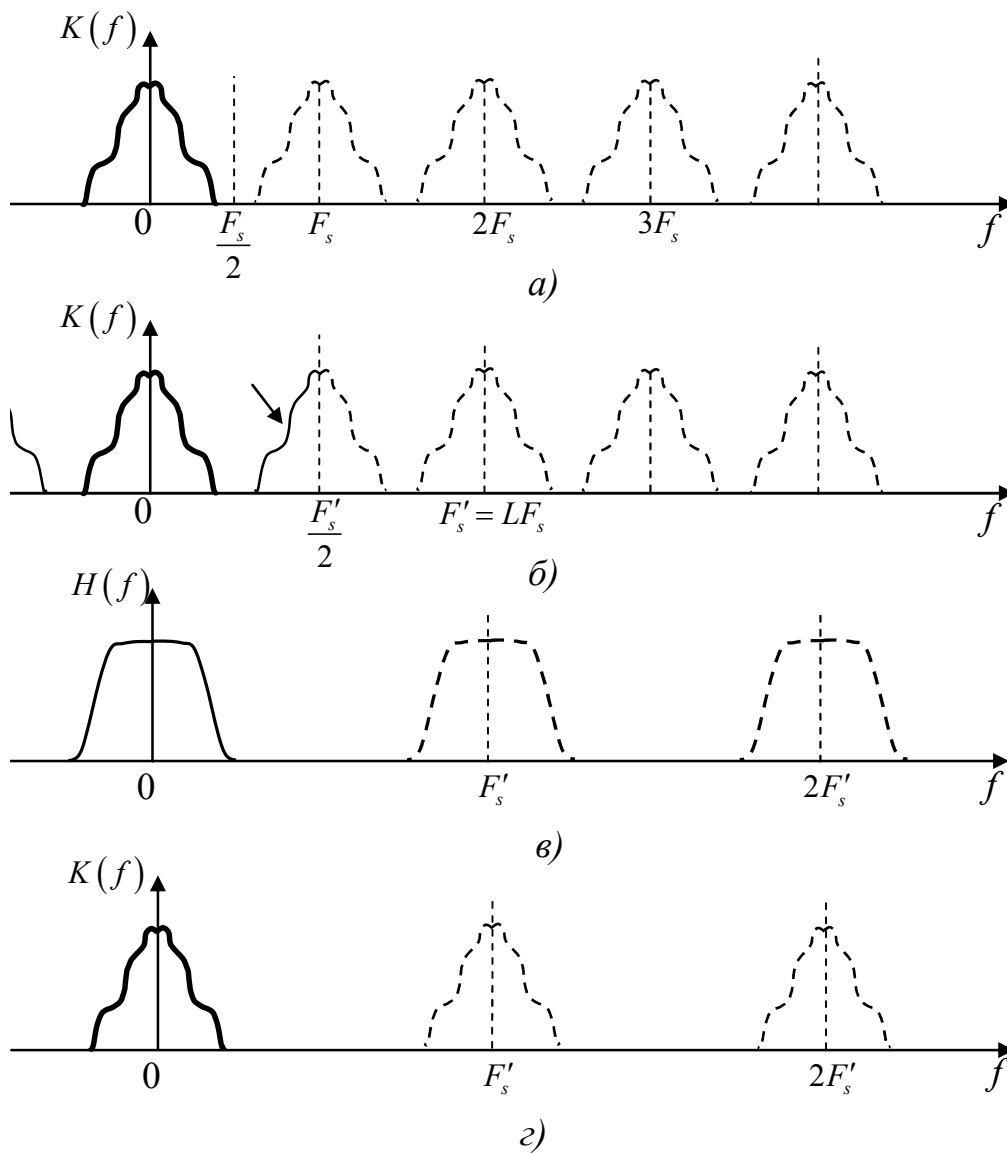


Рисунок 5.15 – а) амплітудний спектр початкового дискретного сигналу, дискретизованого з частотою F_s , б) спектр сигналу після підвищення частоти дискретизації, в) АЧХ фільтра для видалення зайвих складових, г) спектр результуючого сигналу, дискретизованого з частотою F'_s

В часовій області робота інтерполятора зображена на рисунку 5.16. Спочатку між відліками початкового сигналу $x[n]$ вставляють необхідну кількість нульових відліків, а потім отриманий сигнал пропускають через ФНЧ. Блок-схема інтерполятора подана на рис. 5.17, вона складається із експандера (збільшувача) частоти дискретизації в L разів та ФНЧ.

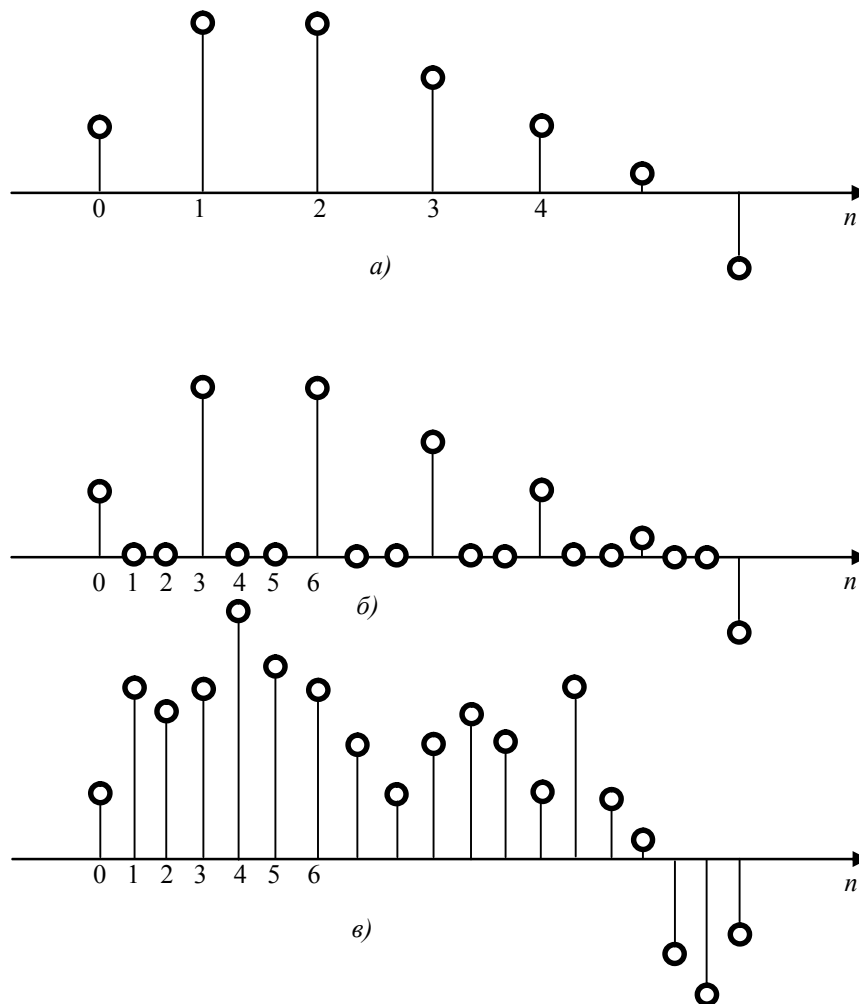


Рисунок 5.16 – Робота інтерполятора: а) початковий дискретний сигнал $x[n]$ з частотою дискретизації F_s , б) сигнал з нульовими відліками, які будуть додані інтерполятором, в) результуючий дискретний сигнал $x'[n]$ з частотою дискретизації $F'_s = LF_s$.

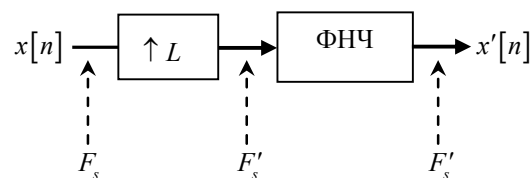


Рисунок 5.17 – Інтерполятор

Введення $L-1$ нулів у сигнал приводить до того, що енергія кожного відліку сигналу розповсюджується по L відлікам вихідного сигналу, тобто кожний відлік

ослаблюється в L разів. Для компенсації цього ефекту можна помножити вихідний сигнал на L .

Зміна частоти дискретизації в неціле число разів

В деяких ситуаціях потрібно змінити частоту дискретизації в неціле число разів, наприклад, в цифровому аудіо, де може знадобитися передача даних з одного пристрою на інший, який підтримує іншу частоту дискретизації. Наприклад, на компакт-дисках використовується частота дискретизації 44.1 кГц, в той час як на студійній цифровій аудіоплівці (digital audio tape, DAT) – 48 кГц. Задача полягає в тому, щоб записати звук, записаний в студії, на компакт-диск для подальшого відтворення. Для виконання цього необхідно зменшити частоту дискретизації в $\frac{48}{44.1}$ разів. Для практичної реалізації такої зміни частоти шукають цілі числа A та B , відношення яких A/B буде максимально близьким до потрібного відношення.

Перетворення частоти виконується в два етапи – інтерполяція даних з кроком A та подальша децимація з кроком B . Необхідно, щоб інтерполяція передувала децимації, оскільки інакше при децимації зникнуть деякі необхідні частотні компоненти. В розглянутому прикладі проводять інтерполяцію з кроком $A=160$, а потім децимацію з кроком 147, тобто на початку частота дискретизації збільшується в 160 разів, а потім зменшується в 147 разів.

На рисунку 5.18а подано блок-схему системи зміни частоти дискретизації з нецілим кроком. Оскільки обидва ФНЧ поєднані послідовно та працюють з сигналом однакової частоти дискретизації, то їх можна поєднати в один фільтр, тоді блок схема матиме вигляд на рис. 5.18б.

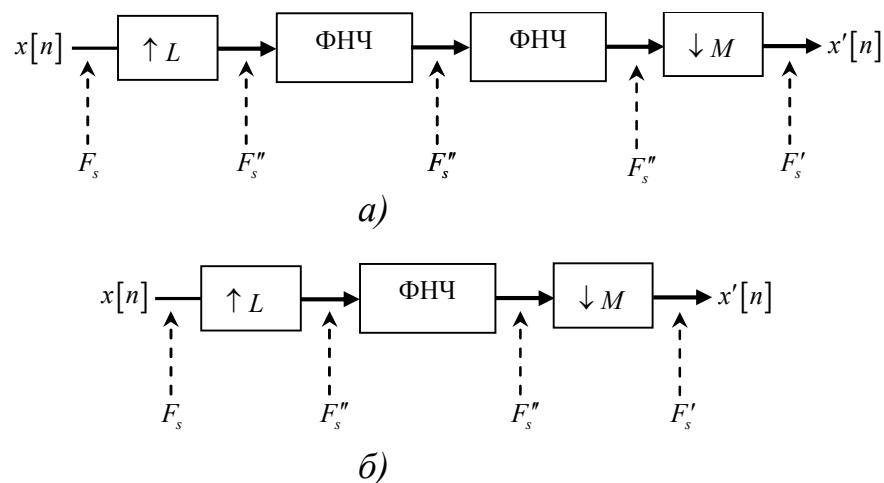


Рисунок 5.18 – Система зміни частоти дискретизації в неціле число разів, повна схема (а) та схема з одним ФНЧ (б)

Як правило, розробка однокаскадних перетворювачів частоти дискретизації неможлива, особливо коли крок перетворення дуже великий. Це пов'язано з тим, що в цьому випадку ФНЧ матиме дуже високий порядок і будуть великі похибки апроксимації. В таких випадках використовують альтернативні підходи: багатокаскадне перетворення частоти (найбільш розповсюджене) або використання багатofазних цифрових фільтрів.

Поняття багатошвидкісної обробки також використовують при сигма-дельта аналого-цифровому перетворенні.

5.8. Фільтр Вінера

Фільтр Вінера – це лінійний фільтр, який використовується при оцінці характеристик та перетворення сигналу, в якому присутня перешкода. Частіше за все фільтр Вінера використовується для виділення корисного сигналу від перешкоди та придушення (фільтрації) перешкоди.

Нехай на вхід системи з передавальною функцією $W_3(p)$ поступає сумарний сигнал $U(t) = G(t) + F(t)$, що складається з корисного сигналу $G(t)$ та перешкоди $F(t)$. Нехай стоїть задача перетворити корисний сигнал за деяким алгоритмом, як описується оператором $H(p)$. В загальному випадку система має якомога точніше відтворити на своєму виході результат перетворення сигналу $G(t)$ відповідно до $H(p)$:

$$Z(t) = H(p)G(t).$$

Застосування оператора $H(p)$ розповсюджує цю задачу не тільки на системи, де треба відтворити на виході корисний сигнал (коли $Z(t) = G(t)$ тобто $H(p) = 1$), але і на інші більш складні класи систем, які виконують різноманітні перетворення вхідного сигналу при наявності шуму. Залежно від виду оператора $H(p)$ говорять про задачі відтворення ($H(p) = \text{const}$), передбачення чи екстраполяції ($H(p) = e^{p\tau}$), інтегрування ($H(p) = \frac{1}{p}$), диференціювання ($H(p) = p$), та інші задачі. В загальному випадку оператор $H(p)$ може бути довільним та описувати потрібне досліднику перетворення вхідного сигналу в системі, що розробляється.

Оскільки присутні перешкоди $F(t)$ та динамічні похибки системи, ідеальне перетворення неможливе, а отже вихідний сигнал системи буде відрізнятися від вхідного на величину випадкової похибки:

$$e(t) = Z(t) - X(t).$$

Задача синтезу таких систем полягає у визначенні таких її характеристик, які найкраще задовольняють деякому критерію оптимальності. Найчастіше використовується критерій мінімуму середньоквадратичної похибки. В цьому випадку задача полягає у визначенні оптимальної передавальної функції системи $W(p)$, при якій значення середньоквадратичної відхилення між сигналами $Z(t)$ та $X(t)$ буде мінімальним.

Залежно від вигляду спектральної густини потужності корисного сигналу та перешкоди, задача пошуку оптимальної системи вирішується по-різному.

В найпростішому випадку, коли спектри сигналу $S_g(\omega)$ та перешкоди $S_f(\omega)$ не накладаються, АЧХ системи обирають відповідно до вимог, що наводилися в попередніх підрозділах. А саме, необхідно, щоб смуга пропускання фільтра була достатньо широкою, щоб була можливість не спотворити спектр корисного сигналу. Водночас, треба щоб спектр перешкоди потрапляв в смугу затримки.

Інший випадок, коли спектр корисного сигналу лежить на нижніх частотах, а спектр шуму близький до білого шуму (має однакову спектральну густину потужності на будь-яких частотах). При цьому треба обирати АЧХ системи так, щоб на нижніх частотах, де сконцентрована енергія корисного сигналу, форма АЧХ була якомога ближчою до форми спектральної густини потужності корисного сигналу $S_g(\omega)$. На інших частотах АЧХ системи повинна швидко спадати відповідно до спаду $S_g(\omega)$.

В найскладнішому випадку, коли спектри сигналу та перешкоди накладаються (лежать на одних і тих самих частотах), задача побудови оптимальної передавальної функції системи стає нетривіальною. Можна показати, що в разі, коли сигнал і перешкода є стаціонарними випадковими процесами з середнім значенням, рівним нулю, то оптимальна передавальна функція визначається в частотній області за виразом:

$$W_{opt}(j\omega) = \frac{S_g(\omega)}{S_g(\omega) + S_f(\omega)} H(j\omega).$$

Отже, єдиними характеристиками, потрібними для задавання передавальної функції, є спектральні густини корисного сигналу та перешкоди, а також спектральна характеристика того перетворення, яке необхідно виконати з сигналом. На попередній вираз також накладаються додаткові вимоги фізичної реалізованості системи. При цьому виділяють ту частину $W_{opt}(j\omega)$, яка є фізично реалізованою, тобто в якій полюси знаходяться в верхній півплощині коренів.

Для цього в ході операції факторизації (запропоновано Г. Боде та К. Шенноном), $W_{opt}(j\omega)$ знаменник попереднього виразу розкладають на комплексні множники:

$$S_g(\omega) + S_f(\omega) = |\Psi(j\omega)|^2 = \Psi(j\omega)\Psi(-j\omega).$$

тут $\Psi(j\omega)$ – функція, всі нулі та полюси якої лежать у верхній півплощині комплексного змінного $j\omega$, тобто яка відповідає фізично реалізованій частині системи;

$\Psi(-j\omega)$ – функція, комплексно спряжена до $\Psi(j\omega)$, всі нулі та полюси якої лежать у нижній півплощині.

У випадку, коли на систему діють корисний сигнал та перешкода, які є *корельованими* між собою стаціонарними випадковими процесами з нульовими середніми значеннями, ставиться задача побудови *фільтра Вінера*.

Імпульсна характеристика $h_{opt}(t)$ фільтра, який задовольняє вимогу фізичної реалізованості (казуальності), повинна задовольняти рівнянню Вінера-Хопфа:

$$\int_{-\infty}^{+\infty} h_{opt}(t) R_u(\tau - t) dt - R_{zu}(\tau) = 0, \quad \tau \geq 0.$$

де $R_u(\tau) = R_g(\tau) + R_f(\tau) + R_{gf}(\tau) + R_{fg}(\tau)$ – кореляційна функція сумарного вхідного сигналу;

$$R_{zu}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T z(t) u(t + \tau) dt - \text{взаємна кореляційна функція вхідного сигналу}$$

$U(t)$ та вихідного сигналу $Z(t)$.

В результаті розв'язку рівняння Вінера-Хопфа Вінером було запропоновано загальний вираз для оптимальної передавальної функції системи (*оптимального фільтра Вінера*):

$$W_{opt}(j\omega) = \frac{1}{2\pi\Psi(j\omega)} \int_0^\infty e^{-j\omega t} dt \int_{-\infty}^\infty \frac{S_{zu}(j\omega)}{\Psi(-j\omega)} e^{j\omega t} d\omega,$$

де $S_{zu}(j\omega) = S_{zg}(j\omega) + S_{zf}(j\omega)$ – взаємна спектральна густина вихідного сигналу $Z(t)$ та сумарного вхідного сигналу $U(t)$;

$$\Psi(j\omega)\Psi(-j\omega) = |\Psi(j\omega)|^2 = S_u(\omega) = S_g(\omega) + S_f(\omega) + S_{gf}(\omega) + S_{fg}(\omega).$$

Варто відмітити, що вразі, коли корисний вхідний сигнал та перешкода некорельовані, то вираз помітно спрощується, оскільки взаємні спектральні густини рівні нулю:

$$S_{gf}(j\omega) = S_{fg}(j\omega) = S_{zf}(j\omega) = 0.$$

На основі цього загального виразу можна отримати оптимальні передавальні функції систем, які виконують різноманітні операції над сигналами: відтворення корисного сигналу, передбачення значень сигналу, диференціювання та інші лінійні перетворення сигналу.

5.9. Підсумки до розділу 5 та рекомендована література

В цьому розділі була розглянута задача обробки сигналів з метою позбавлення їх від небажаних складових. Це виконується за допомогою лінійних фільтрів, які проводять обробку таким чином, що спектральний склад вихідного сигналу фільтра буде відрізнятися від спектра вхідного сигналу. Як саме фільтр змінить сигнал залежить від того, який його тип, тобто від того, на яких частотах коефіцієнт передачі фільтра має великі значення, а на яких він близький до нуля. Було розглянуто основні параметри та характеристики фільтрів, такі як частота зрізу, частота затримки, коефіцієнт передачі, дано поняття смуги пропускання, затримки та перехідної смуги.

Коротко було обговорено основні етапи проектування фільтрів, і більш докладно розказано про задачу апроксимації фільтрів, тобто знаходження математичного опису АЧХ реального фільтра, який за своїми властивостями найбільш близький до властивостей потрібного ідеального фільтра.

Також були стисло охарактеризовані основні шляхи проектування цифрових фільтрів на основі аналогових прототипів. Вказано на відмінності між фільтрами з кінцевою та безкінечною імпульсними характеристиками.

Багато уваги в цьому розділі було приділено питанню роботи систем багатошвидкісної обробки сигналів, при якій необхідно змінювати частоту дискретизації дискретних сигналів. Докладно розглянуто шляхи децимації та інтерполяції, та на основі аналізу спектрального представлення дискретизації було обґрунтовано побудову інтерполятора і дециматора.

Питання фільтрації, хоча напряду і не пов'язані з теорією аналізу сигналів, але є дуже важливими, оскільки фільтрація передуює аналізу і проводиться для того, щоб очистити аналізований сигнал від шумів та інших небажаних складових. Від того, наскільки якісно буде виконана фільтрація, залежить те, наскільки результативною буде подальший аналіз.

В цьому розділі було лише побіжно викладено основні відомості про фільтрацію, необхідні для розуміння її мети та особливостей застосування. Цифрова фільтрація – достатньо добре розроблена галузь обробки сигналів. Класичними підручниками з проектування цифрових фільтрів є книги Хеммінга [31] та Богнера і Константінідіса [32]. Також можна рекомендувати класичний підручник з цифрової обробки сигналів [2].

5.8. Питання для самоперевірки

1. Навіщо потрібна фільтрація і в яких випадках її треба застосовувати?
2. Які існують типи фільтрів, як за виглядом АЧХ визначити його тип? Якими основними параметрами описується АЧХ фільтра?
3. Які етапи треба виконати для проектування аналогового та цифрового фільтра? В чому відмінність і схожість кожного з цих етапів для різних фільтрів?
4. Що таке «апроксимація фільтра» і якими методами вона виконується?
5. Чим відрізняються БІХ- та КІХ-фільтри і які існують підходи до їх проектування?
6. Що відображає структура фільтра і як її побудувати?
7. В чому полягає багатошвидкісна обробка сигналів, якими операціями вона проводиться? Що відбувається зі спектром сигналу при інтерполяції та децимації?

6. АНАЛІЗ ВИПАДКОВИХ СИГНАЛІВ

Equation Chapter 1 Section 6

6.1. Випадкові сигнали

Всі методи аналізу сигналів, які були викладені раніше в цьому посібнику, були орієнтовані на те, що аналізований сигнал збережений в деякому вигляді в пам'яті комп'ютера. Іншими словами, весь сигнал нам відомий – ми знаємо або можемо розрахувати всі його значення в будь-який момент часу. Ми неявно мали на увазі, що аналізовані сигнали є детермінованими – якби ми змогли повторити процес вимірювання сигналу ще раз, ми б отримали такий саме сигнал.

В цьому розділі йтиме мова про інший клас сигналів – випадкові сигнали. На відміну від детермінованого сигналу, значення якого ми знаємо точно, значення випадкових сигналів в кожний момент часу заздалегідь невідомі та можуть бути передбачені лише з певною імовірністю, меншою одиниці. Характеристики таких сигналів є статистичними та мають імовірнісний вид.

Наведемо приклад експерименту, в якому отримаємо випадковий сигнал. Нехай ми досліджуємо шумові характеристики резистора та вимірюємо залежність напруги на ньому від часу $U(t)$ при попусканні через нього постійного струму. Визначивши умови експерименту (температуру, ланцюг вимірювання, параметри струму, який пропускають через резистор, систему вимірювання та реєстрації вихідного сигналу), ми можемо зняти деяку залежність $U_1(t)$. Через те, що в резисторі при проходженні через нього струму буде відбуватися багато процесів, що є випадковими за своєю природою (теплові коливання решітки, дробовий шум та ін.), напруга $U_1(t)$ матиме випадковий характер. Ця напруга характеризує всі процеси, що відбуваються в резисторі при проходженні через нього струму.

Уявімо собі, що наступного дня ми провели той самий експеримент – виміряли напругу $U_2(t)$ на резисторі, причому всі умови зберегли такими самими, як і в попередньому разі. Не дивлячись на це, $U_2(t)$ буде відрізнятися від $U_1(t)$ внаслідок того, що процеси, які відбуваються в резисторі, носять випадковий характер, і вони не повторяться ще раз. Значення $U_1(t)$ та $U_2(t)$ будуть різними в одні і ті самі моменти часу, не дивлячись на те, що ці дві напруги характеризують одне і те саме явище. Так само, якщо повторити дослід ще раз, і виміряти сигнал $U_3(t)$ – він також буде відмінний від двох попередніх, але так само буде характеризувати шум резистора. Скільки б разів ми не повторювали дослід, кожного разу будемо отримувати інший сигнал. Для того, щоб отримати інформацію про те, якими насправді є характеристики шуму $U(t)$, нам потрібні всі виміряні сигнали. Лиша маючи їх і провівши усереднення, можна визначити, якими насправді є шумові характеристики резистора.

Шумова напруга на резисторі – це приклад випадкового сигналу. Напруги, які ми вимірювали в кожному досліді – це окремі *реалізації* цього випадкового сигналу.

Випадковою функцією $X(t)$ називають таку функцію свого аргументу, значення якої при будь-якому значенні t є випадковою величиною.

Якщо аргумент випадкової функції може приймати *будь-які* значення на заданому інтервалі (скінченному або безкінечному), то $X(t)$ називають *випадковим процесом*.

Якщо аргумент випадкової функції може приймати тільки *певні* значення на заданому інтервалі (скінченному або безкінечному), то $X(t)$ називають *випадковою послідовністю*.

Випадковий процес – це математична модель неперервного випадкового сигналу, а випадкова послідовність – математична модель дискретного випадкового сигналу.

Нехай розглядаємо випадкову функцію $X(t)$ та проведемо для її дослідження n незалежних дослідів. В результаті кожного дослідів буде отримана одна з можливих *реалізацій* $x_1(t), x_2(t), \dots, x_n(t)$ випадкової функції $X(t)$. В нашому досліді ми отримаємо набір кривих (рис. 6.1), кожна з яких певною мірою характеризує властивості випадкової функції $X(t)$. Разом вони називаються *ансамблем реалізацій* випадкового процесу.

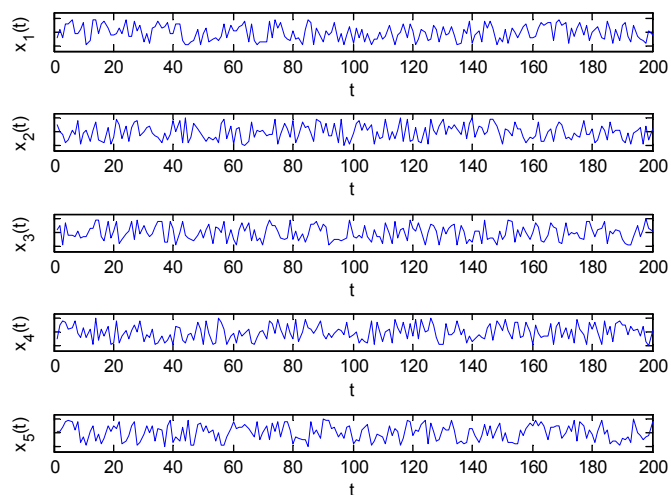


Рисунок 6.1 – Ансамбль реалізацій випадкового процесу $X(t)$

Говорять, що *до реєстрації* випадковий сигнал необхідно розглядати як випадковий процес, що є ансамблем реалізацій – сукупністю функцій часу $x_i(t)$, які підкоряються спільній для них статистичній закономірності.

Одна з цих функцій, яка стала повністю відомою після вимірювання (реєстрації), називається *реалізацією* випадкового процесу. Після того, як реалізація зареєстрована та збережена в пам'яті комп'ютера, вона може вважатися вже не

випадковою, а детермінованою функцією часу, оскільки досліднику вона відома точно. В той же час, неможливо передбачити, якою саме буде конкретна реалізація випадкового процесу в кожному досліді. Але тим не менш, вона характеризує випадковий процес.

Якщо зареєстрований деякий сигнал, я необхідно використати для його аналізу математичний апарат випадкових процесів, то можна розглядати цей детермінований сигнал як випадковий сигнал, який має лише одну єдину реалізацію.

6.2. Основні характеристики випадкових сигналів

Нехай маємо всі реалізації $x_i(t)$ випадкового сигналу $X(t)$. Зафіксуємо одне зі значень аргументу при $t = t_1$. Кожна реалізація є випадковою, отже, і значення кожної реалізації в певний момент часу також є випадковою величиною. Очевидно, що якщо ми б хотіли знати значення випадкового сигналу в момент часу t_1 , то нам було б достатньо знати закон розподілу такої випадкової величини:

$$X_1 = X(t_1).$$

Одномірна інтегральна функція розподілу ймовірностей визначає імовірність P того, що значення випадкового сигналу не перевищують деяке значення x :

$$F(X|t_1) = P(X(t_1) < x). \quad (6.1)$$

$X(t_1)$ буде неперервною випадковою величиною.

Одномірна густина ймовірностей визначає імовірність потрапляння значень випадкового сигналу в малий інтервал:

$$p(X|t_1) = P(x < X(t_1) < x + \Delta x) \quad (6.2)$$

З теорії ймовірностей відомо, що густина ймовірностей невід'ємна, а також виконуються такі рівності:

$$p(X|t_1) > 0, \quad \int_{-\infty}^{\infty} p(X|t_1) dx = 1, \quad p(X|t_1) = \frac{dF(X|t_1)}{dx}.$$

Характеристична функція для випадкового сигналу визначається за виразом:

$$\theta(u) = \int_{-\infty}^{\infty} p(X|t_1) e^{-jux} dx \quad (6.3)$$

В загальному випадку, коли імовірнісні характеристики (6.1)-(6.3) є функціями часу, випадковий сигнал називають нестационарним.

В багатьох практичних застосуваннях достатньо знати лише одновимірну густину імовірності. Наприклад, якщо $X(t)$ є величиною відхилення снаряду від траєкторії в залежності від часу польоту, то обравши як t_1 час польоту снаряду до цілі, за допомогою густини ймовірностей $p(X|t_1)$ ми матимемо можливість характеризувати закон відхилення снаряду від цілі.

Але одновимірна густина імовірності не може слугувати повною характеристикою випадкового сигналу, оскільки не відображає взаємну залежність значень випадкового сигналу в різні моменти часу. Ця характеристика також може бути корисною для описання та вивчення властивостей випадкового сигналу. Для отримання такої характеристики випадкового сигналу можна обрати два значення аргументу t_1 та t_2 . Відповідні значення сигналу $X(t_1)$ та $X(t_2)$ будуть випадковими величинами X_1 та X_2 , які разом можуть бути охарактеризовані відповідною двовимірною густиною імовірності $p(X_1, X_2 | t_1, t_2)$. Якщо для аргументу зафіксувати три значення та розглянути тривимірну густину імовірності, то ми отримаємо ще більш докладну характеристику випадкового сигналу і так далі.

Випадковий сигнал можна вважати повністю заданим, якщо задані всі багатовимірні закони розподілу для будь-яких значень t_1, t_2, \dots, t_n .

Це не завжди зручно, тому часто замість самих законів розподілу задають їх параметри, найбільш зручними з яких є *початкові* $m_{j_1 j_2 \dots j_n}$ та *центральні* $\mu_{j_1 j_2 \dots j_n}$ *моменти* різних порядків, тобто математичні сподівання добутоків відповідних степенів значень випадкового сигналу:

$$m_{j_1 j_2 \dots j_n} = E[X^{j_1}(t_1) X^{j_2}(t_2) \dots X^{j_n}(t_n)],$$

$$\mu_{j_1 j_2 \dots j_n} = E\left[\left(X(t_1) - E[X(t_1)]\right)^{j_1} \left(X(t_2) - E[X(t_2)]\right)^{j_2} \dots \left(X(t_n) - E[X(t_n)]\right)^{j_n}\right],$$

де *порядком моменту* називається сума $j_1 + j_2 + \dots + j_n$.

З нескінченного числа моментів найбільш важливим для більшості задач є моменти першого та другого порядків.

Початковий момент першого порядку називається *математичним сподіванням* (mean value) випадкової величини

$$m_1 = \overline{X(t_1)} = E[X(t_1)] = \int_{-\infty}^{+\infty} X(t_1) p(X|t_1) dx.$$

Математичне сподівання характеризує середнє значення випадкового сигналу в певний момент часу t_1 . Випадковий сигнал, який має нульове математичне сподівання, називається *центрованим*.

Початкові моменти другого порядку можуть бути двох типів: моменти другого порядку з одною ординатою випадкової функції:

$$m_2 = E[X^2(t_1)],$$

або змішані моменти другого порядку:

$$m_{1,1} = E[X(t_1) X(t_2)].$$

Значення $m_{1,1}$ називається *коефіцієнтом коваріації* випадкового сигналу – він є середнім добутком значень випадкового сигналу в моменти часу t_1 та t_2 . Цей

коефіцієнт в загальному випадку залежить не тільки від взаємного розташування моментів часу t_1 та t_2 , а також і від відстані між ними; в цьому випадку говорять про коваріаційну функцію.

Часто також застосовують центральний момент другого порядку з одною ординатою, який називається *дисперсією* (variance) випадкової величини:

$$D[X(t)] = E\left[(X(t) - \bar{x})^2\right].$$

Для сигналів, які мають електричну природу, дисперсія характеризує *середню потужність відхилень* випадкового сигналу від його середнього значення. Відхилення та коливання значень сигналу навколо середнього називаються також флуктуаціями. Середню величину цих флуктуацій для неперервного випадкового сигналу можна отримати за такими виразами:

$$D = \sigma^2 = \int_{-\infty}^{\infty} (X(t) - m_1)^2 p(X|t_1) dx = \int_{-\infty}^{\infty} X^2(t) p(X|t_1) dx - m_1^2.$$

Середньоквадратичне відхилення (standard deviation) – квадратний корінь із дисперсії, є мірою *амплітуди* розкиду значень випадкового сигналу навколо його середнього значення.

$$\sigma = \sqrt{D}.$$

Ще один змішаний центральний момент другого порядку називають *кореляційною функцією*:

$$\begin{aligned} R_{XX}(t_1, t_2) &= E\left[(X(t_1) - \bar{x})(X(t_2) - \bar{x})\right] = \\ &= m_{1,1} - m_1(t_1)m_1(t_2). \end{aligned}$$

Ця функція характеризує ступінь статистичного зв'язку між відхиленнями значень одного і того самого випадкового сигналу від його середнього значення, які можна спостерігати в моменти часу t_1 та t_2 . Оскільки досліджується один випадковий сигнал, то така кореляційна функція називається *автокореляційною функцією* (АКФ).

Якщо аналізуються два випадкові сигнали $X(t)$ та $Y(t)$, і необхідно знайти статистичний зв'язок між ними в моменти часу t_1 та t_2 , то можна розрахувати *функцію взаємної кореляції* (взаємнокореляційну функцію, ВКФ):

$$\begin{aligned} R_{XY}(t_1, t_2) &= E\left[(X(t_1) - \bar{X})(Y(t_2) - \bar{Y})\right] = \\ &= m_{1,1} - m_{X1}(t_1)m_{Y1}(t_2). \end{aligned}$$

Слід зауважити, що наявність статистичного зв'язку між сигналами (ненульові значення ВКФ) не значить, що процеси, які досліджуються, дійсно зв'язані в реальному житті. Кореляція не свідчить про каузальність, тобто причинно-наслідкові зв'язки між процесами або об'єктами. Вона лише значить, що формули, які використовуються, дають ненульовий результат розрахунків, і процеси статистично зв'язані. Для того, щоб визначити, чи дійсно кореляція показує на зв'язок між процесами, треба залучити додаткову інформацію про природу сигналів,

спланувати та провести додаткові експерименти, або провести тести на сурогатних даних.

Розділ теорії випадкових функцій, який має справу тільки з моментами перших двох порядків, називається кореляційною теорією випадкових функцій.

6.2. Класифікація випадкових сигналів

З метою розробки більш ефективних методів аналізу, що пристосовані до окремих видів випадкових сигналів, порівняно з методами, які є більш загальними, необхідним є дослідження можливої класифікації випадкових сигналів за певними ознаками.

По залежності ймовірнісних характеристик від початку відліку часу.

Стаціонарні випадкові сигнали – для них характерна незалежність від початку відліку часу таких ймовірнісних характеристик, як моменти, функції розподілу та т.ін. Багатовимірні закони розподілу для таких сигналів залежать лише від взаємного розташування моментів часу t_1, t_2, \dots, t_n , а не від самих значень цих величин, тобто

$$p(x_1, x_2, \dots, x_n | t_1, t_2, \dots, t_n) = p(x_1, x_2, \dots, x_n | t_1 + \tau, t_2 + \tau, \dots, t_n + \tau)$$

для довільного значення τ .

Виділяють стаціонарність в широкому та вузькому смислі:

Стаціонарний в широкому смислі сигнал – для якого математичне сподівання та дисперсія постійні (не залежать від часу), а автокореляційна функція $R(t_1, t_2)$ залежить лише від відстані в часі між двома значеннями сигналу t_1 та t_2 .

Стаціонарний у вузькому смислі сигнал – всі його скінченновимірні розподіли ймовірностей не змінюються при зсуві сигналу по вісі часу.

Сутність стаціонарності полягає в однорідності (незмінності) умов, які обумовлюють характер зміни сигналу в часі; на практиці такі сигнали зустрічаються достатньо рідко.

Нестационарні випадкові сигнали – це сигнали для яких ймовірнісні характеристики змінюються залежно від того, в який момент часу вони розраховуються. Це найбільш широкий клас процесів, які найчастіше зустрічаються на практиці, зокрема, при дослідженні сигналів біологічних об'єктів.

За областями існування випадкових сигналів.

Виділяють дискретні та неперервні випадкові сигнали. Дискретні – які існують лише у певні моменти часу, неперервні – для яких можна отримати значення у будь-який момент часу.

По наявності залежності значень сигналу від попередньої історії.

Якщо поведінка випадкового процесу в наступні моменти часу не залежить від поведінки в попередні моменти часу, а визначається тільки значенням в поточний момент часу, то такі процеси називаються марковськими процесами. Всі інші процеси – немарковські.

За наявністю зв'язків між середніми по часу та середньому по ансамблю.

Для *ергодичних* процесів характерна рівність (в граничному сенсі) параметрів, які отримані осередненням за ансамблем реалізацій, та осередненням за часом. Якщо процес *неергодичний*, то ймовірнісні характеристики, які отримані при осередненні за реалізаціями, та які отримані при осередненні в часі (лише по одній реалізації), не рівні.

Найбільш зручним для аналізу сигналів є випадок, коли аналізований випадковий сигнал є ергодичним – тоді не треба повторювати дослід кілька разів, щоб отримати багато реалізацій. У випадку ергодичного сигналу параметри, отримані по багатьом реалізаціям будуть такими самими, які ми зможемо отримати по одній з реалізацій.

Також бажаним є випадок стаціонарного сигналу, оскільки в цьому разі неважливий момент часу, в який ми починаємо розраховувати параметри сигналу. Варто зауважити, що часто гіпотези ергодичності та стаціонарності аналізованого приймаються дослідником заздалегідь, без доведення. Але існують методи дослідження стаціонарності та перевірки ергодичності.

За типами законів розподілу значень випадкового сигналу.

Випадкові сигнали бувають гаусівські, релеївські, пуасонівські та ін.

За властивостями приростів випадкового сигналу.

Приріст сигналу – це різниця між його наступними та попередніми значеннями. Сигнали бувають зі стаціонарними приростами, з незалежними приростами, з некорельованими приростами, з ортогональними приростами.

6.3. Поняття про оцінки характеристик випадкових сигналів

У загальному випадку випадковий сигнал є сукупністю можливих реалізацій, отже охарактеризувати його треба, аналізуючи всі реалізації. Оскільки нескінченна кількість реалізацій недоступна досліднику, завжди обирають кілька реалізацій та розраховують параметри випадкового сигналу, ґрунтуючись лише на них. Через це в результаті розрахунків параметрів реальних випадкових сигналів будемо мати не точне значення деякого параметра сигналу, а лише деяке наближене значення – його *оцінку*. Розглянемо коротко основні поняття теорії оцінок.

Якщо проводиться оцінювання деякого параметра α , то *зміщення* (відхилення) визначається як різниця істинного значення (яке було б взяте із всієї генеральної сукупності), та його оцінки $M[\alpha]$:

$$\Delta = \alpha - M[\alpha].$$

Якщо зміщення дорівнює нулю, то оцінка співпадає з істинним значенням; якщо зміщення не дорівнює нулю, то воно представляє похибку визначення α і оцінене значення $M[\alpha]$ називається *зміщеним*.

Дисперсія величини α характеризує розкид значень вимірюваного параметра навколо середнього значення. Якщо оцінка є точною, то середнє значення повинне бути рівним істинному значенню, а розкид має бути все меншим при збільшенні кількості дослідів.

Можна сказати, що дисперсія величини α є мірою ширини максимуму функції розподілу густини ймовірностей значень α . При невеликій дисперсії пік функції достатньо вузький, а при дисперсії, яка прямує до нуля, оцінене значення прямує до істинного (в разі, коли оцінка незміщена).

Для того, щоб отримати оцінку α , проводять розрахунки для кількох реалізацій, результати яких потім усереднюють. Якщо дисперсія оцінки величини α прямує до нуля при збільшенні кількості реалізацій, за якими проводиться усереднення, то оцінка називається *змістовною*. Якщо оцінка не змістовна, то при збільшенні реалізацій дисперсія не буде зменшуватися, а розрахунок не буде уточнюватися.

Отже, необхідно, щоб всі оцінки параметрів сигналів були незміщеними та змістовними.

6.4. Неперервні стаціонарні ергодичні випадкові сигнали

На практиці дуже часто отримання характеристик випадкового сигналу приходить робити лише по одній реалізації скінченної тривалості, і через це приймається припущення про ергодичність та стаціонарність аналізованого сигналу.

Для стаціонарного в широкому сенсі сигналу математичне очікування та дисперсія не залежать від часу, а кореляційна функція залежить не від самих моментів часу, а тільки від інтервалу між ними:

$$R(t_1, t_2) = R(t_2 - t_1) = R(\tau).$$

Кореляційна функція стаціонарного випадкового процесу є парною: $R(\tau) = R(-\tau)$, її абсолютні значення не перевищують її значення при $\tau = 0$.

Часто для аналізу використовують *нормовану кореляційну функцію*:

$$r(\tau) = \frac{R(\tau)}{D}.$$

Функції $R(\tau)$ та $r(\tau)$ характеризують зв'язок (кореляцію) між значеннями стаціонарного сигналу $X(t)$, які розділені проміжком тривалості τ . Чим повільніше спадають ці функції зі зростанням абсолютного значення τ , тим більшим є проміжок, протягом якого існує статистичний зв'язок між миттєвими значеннями випадкового процесу, і тим повільніше змінюються в часі його реалізації.

Для стаціонарних ергодичних процесів осереднення по одній реалізації статистично еквівалентне осередненню по всьому ансамблю.

Запишемо основні статистичні характеристики для стаціонарних ергодичних випадкових сигналів тривалістю T у неперервному випадку:

$$m_{x1} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt,$$

$$D_x = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} (x(t) - m_{x1})^2 dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t) dt - m_{x1}^2,$$

$$R(\tau) = E[(x(t) - m_{x1}) \cdot (x(t - \tau) - m_{x1})] = E[x(t)x(t - \tau)] - m_{x1}^2 =$$

$$= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t - \tau) dt - m_{x1}^2.$$

6.5. Дискретні стаціонарні ергодичні випадкові сигнали

Надалі будемо розглядати лише дискретні стаціонарні ергодичні сигнали, оскільки вони є найбільш розповсюдженими на практиці.

Для дискретного випадку математичне сподівання випадкової ергодичної величини $x[n]$ (або середнє значення) визначається за виразом:

$$M[x[n]] = \lim_{N \rightarrow \infty} \left\{ \frac{1}{2N+1} \sum_{n=-N}^N x[n] \right\}.$$

Автокореляційна функція (АКФ) в загальному випадку визначається так:

$$r_{xx}[m] = E[x[n+m]x^*[n]] = \lim_{N \rightarrow \infty} \left\{ \frac{1}{2N+1} \sum_{n=-N}^N x[n+m]x^*[n] \right\}.$$

На практиці ми маємо лише послідовність обмеженої тривалості, а отже отримаємо не саму функцію, а лише її оцінку:

$$\hat{r}_{xx}[m] = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x[n+m]x^*[n]. \quad (6.4)$$

Можна показати, що ця оцінка є незміщеною, оскільки її математичне сподівання дорівнюватиме істинному значенню АКФ. Часто також використовують зміщену оцінку АКФ:

$$\tilde{r}_{xx}[m] = \frac{1}{N} \sum_{n=0}^{N-m-1} x[n+m]x^*[n]. \quad (6.5)$$

Обидві оцінки задовольняють співвідношення:

$$r_{xx}[m] = r_{xx}^*[-m], \quad (6.6)$$

та зберігають повну потужність сигналу:

$$\tilde{r}_{xx}[0] = \hat{r}_{xx}[0] = \frac{1}{N} \sum_{n=0}^{N-1} |x[n]|^2.$$

При аналізі двох випадкових сигналів вводять поняття *взаємнокореляційної функції* (ВКФ):

$$\begin{aligned} r_{xy}[m] &= M\{x[n+m]y^*[n]\} = \lim_{M \rightarrow \infty} \left\{ \frac{1}{2M+1} \sum_{n=-M}^M x[n+m]y^*[n] \right\}, \\ \hat{r}_{xy}[m] &= \frac{1}{N-m} \sum_{n=0}^{N-m-1} x[n+m]y^*[n], \\ \tilde{r}_{xy}[m] &= \frac{1}{N} \sum_{n=0}^{N-m-1} x[n+m]y^*[n]. \end{aligned} \quad (6.7)$$

В загальному випадку співвідношення (6.6) не виконується: $r_{xy}[m] \neq r_{xy}^*[-m]$, але виконується співвідношення $r_{xy}[m] = r_{yx}^*[-m]$

6.6. Спектральний аналіз дискретних стаціонарних ергодичних випадкових сигналів

Одною з основних задач в аналізі випадкових сигналів, як і для детермінованих, є спектральний аналіз за Фур'є. Для випадкових сигналів буде отримано не спектр сигналу (як у випадку детермінованого сигналу), а *оцінку спектру*. Для випадкових сигналів зазвичай отримують оцінки *спектральної густини потужності* (СГП).

Існують дві групи методів оцінки СГП: параметричні та непараметричні.

Непараметричні методи – в яких для отримання оцінок спектра випадкового сигналу використовують тільки інформацію, яка міститься у відліках сигналу без додаткових припущень. До основних непараметричних (класичних) методів відносяться: корелограмний метод (метод Блекмана і Тьюкі), періодограмні методи Даньєлла, Бартлетта та Уелча.

Параметричні методи – ґрунтуються на підборі *параметрів моделі* часового ряду, яка відповідає аналізованому випадковому сигналу, а визначення оцінки спектра проводиться на основі цих параметрів. На першому етапі параметричного методу проводять вибір параметричної моделі часового ряду. Основні – це авторегресійна (autoregressive, AR); осереднення зі зсувом (moving average, MA), її в деяких джерелах називають моделлю ковзного середнього; комбінована авторегресійна-осереднення зі зсувом (autoregressive-moving average, ARMA). На другому етапі розраховують оцінки параметрів моделі. На третьому – оцінені параметри моделі підставляють у математичний вираз для спектральної густини потужності, який відповідає даній моделі.

Корелограмний метод оцінки спектру

Основою корелограмних методів оцінки СГП є *теорема Вінера-Хінчіна-Колмогорова*: автокореляційна функція та спектральна густина потужності випадкового сигналу пов'язані перетворенням Фур'є:

$$P_{xx}(\omega) = \sum_{n=-\infty}^{+\infty} r_{xx}[n]e^{-j\omega nT_s}, \quad (6.8)$$

де T_s - інтервал дискретизації сигналу.

Сутність корелограмного методу – підстановка в вираз (6.8) скінченної послідовності оцінки АКФ із співвідношень (6.4) або (6.5). Якщо використовувати незміщену оцінку з (6.4) для максимально можливого зсуву L , то можна записати:

$$\hat{P}_{xx}(\omega) = \sum_{n=-L}^L \hat{r}_{xx}[n]e^{-j\omega nT_s}.$$

На практиці L беруть набагато менше довжини послідовності $L \ll N$. ($L \cong N/10$), де N – довжина сигналу. Кореляцію для додатних індексів можна отримати за (6.4), а для від’ємних – з використанням (6.6).

Часто корелограмний метод доповнюють з використанням функцій вікна для зменшення ефекту Гібса:

$$\hat{P}_{xx}(\omega) = \sum_{n=-L}^L \hat{r}_{xx}[n] w[n] e^{-j\omega n T_s},$$

де $w[n]$ – функція вікна.

Оскільки множення в часовій області відповідає згортці спектрів, то вибирати функцію вікна необхідно з тих міркувань, щоб її СГП була достатньо сконцентрованою на частотах в околі нульової частоти.

Вводять також поняття взаємної густини потужності для двох випадкових сигналів:

$$P_{xy}(\omega) = \sum_{n=-\infty}^{\infty} r_{xy}[n] e^{-j\omega n T_s}.$$

Можна вважати, що ця характеристика описує імовірнісний зв’язок між спектрами.

Періодограмні методи оцінки СГП

Періододіаграма.

Ще один класичний метод оцінки СГП ґрунтується на прямому використанні властивості ергодичності процесу

$$P_{xx}(\omega) = \lim_{N \rightarrow \infty} M \left[\frac{1}{(2N+1)T} \left| \sum_{n=-N}^N x[n] e^{-j\omega n T_s} \right|^2 \right].$$

На практиці іноді використовують тільки додатні номери відліків та нехтують операцією осереднення:

$$P_{xx}(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j\omega n T_s} \right|^2.$$

Ця оцінка виявляється незмістовною і використовується дуже рідко, оскільки операція осереднення не виконується, і зі збільшенням числа відліків сигналу значення періододіаграми мають значні флуктуації.

Періододіаграма Даньєлла.

Даньєлл запропонував для згладжування швидких флуктуацій періододіаграми використовувати осереднення по сусіднім частотам спектру. Якщо розрахунки

$P_{xx}(f)$ проводяться на сітці частот $f_k = k \frac{1}{N}$, $0 \leq k \leq N-1$, то модифікована оцінка

періодограми на частоті f_i може бути отримана шляхом осереднення спектральних відліків на m частотах з кожного боку цієї:

$$P_D[f_i] = \frac{1}{2m+1} \sum_{n=i-m}^{i+m} P_{xx}(f_i).$$

Узагальненням такого підходу є обробка вибіркового спектру за допомогою фільтру нижніх частот.

Періодограма Бартлетта.

Згладжування періодограми в цьому методі ґрунтується на створенні псевдоансамблю періодограм за рахунок ділення випадкової послідовності $x[n]$ з N відліків на K сегментів, що не перекриваються, по D відліків в кожному ($DK \leq N$). Кожний m -ий сегмент буде містити відліки

$$x^{(m)}[n] = x[mD + n], \quad 0 \leq n \leq D-1.$$

По кожному сегменту розраховують спектр

$$P_{xx}^{(m)}(f) = \frac{1}{D} \left| \sum_{k=0}^{D-1} x^{(m)}[k] e^{-j2\pi f k T_s} \right|^2,$$

а потім проводять усереднення для всіх частот для отримання періодограми Бартлетта:

$$P_B(f_i) = \frac{1}{K} \sum_{m=0}^{K-1} P_{xx}^{(m)}(f_i).$$

Можна показати, що величина дисперсії оцінки СГП за методом Бартлетта обернено пропорційна кількості сегментів.

Періодограма Уелча.

Метод отримання оцінки СГП за Уелчем ще називають методом осереднення модифікованих періодограм. Уелч модифікував основну схему Бартлетта за рахунок використання віконної функції та використання сегментів, які перекриваються. Перед розрахунком періодограми кожного сегменту він обробляється за допомогою вікна для зменшення бокових пелюстків спектру та зменшення зміщення оцінки СГП. Метою перекриття сегментів є збільшення кількості сегментів, за якими проводять осереднення, щоб таким чином зменшити дисперсію оцінки СГП. Уелч запропонував використовувати перекриття на 50 % та вікно Ханна. Наразі метод Уелча є одним з найбільш популярних періодограмних методів оцінки СГП.

Якщо запис сигналу розділити на K сегментів по D відліків в кожному зі зсувом S відліків між сусідніми сегментами, то кількість сегментів буде рівна

$K = \left\lceil \frac{N-D}{S+1} \right\rceil$ (ціла частина). При зважуванні k -й сегмент буде містити відліки:

$$x^{(k)}[n] = w[n]x[n + kS], \quad 0 \leq n \leq D-1$$

для номерів сегментів $0 \leq k \leq K-1$.

Спектр зваженого сегменту буде визначатися виразом:

$$P_{xx}^{(k)}(f) = \frac{1}{UD} \left| \sum_{n=0}^{D-1} x^{(k)}[n] e^{-j2\pi f n T_s} \right|^2,$$

де

$$U = \sum_{m=0}^{D-1} w^2[m]$$

– енергія вікна, яке використовується. Ділення на U дозволяє усунути ефект впливу енергії вікна на зміщення в оцінці СГП.

Середнє значення періодограм зважених сегментів дає оцінку періодограми Уелча:

$$P_W(f_i) = \frac{1}{K} \sum_{k=0}^{K-1} P_{xx}^{(k)}(f).$$

Параметричні методи оцінки СГП

Параметричні методи ґрунтуються на параметричних моделях випадкових процесів. Найбільш розповсюдженими є методи, які ґрунтуються на *авторегресійній моделі* формування сигналу. В ній вважається, що досліджуваний сигнал є вихідним сигналом фільтра, на вхід якого подають білий шум з відомою дисперсією. При цьому СГП буде деякою функцією параметрів моделі фільтра, а не самого часового ряду. Якщо входом фільтра є білий шум, то СГП вихідного процесу повністю описується за допомогою параметрів моделі та дисперсії шумового процесу. Значення цих параметрів та дисперсії шуму визначаються за автокореляційною функцією аналізованого випадкового сигналу. Оцінка автокореляційної функції повинна бути розрахована на основі виміряних значень випадкового сигналу.

Вихідний сигнал фільтра, який вважається аналізованим випадковим сигналом, можна представити лінійним різницеvim рівнянням:

$$\begin{aligned} x[n] &= \sum_{k=0}^{+\infty} h[k] u[n-k] = \\ &= \sum_{k=0}^q b_k u[n-k] + \sum_{k=1}^p a_k x[n-k] \end{aligned} \quad (6.9)$$

Проста *авторегресійна модель* отримується в разі, коли формуючий фільтр є рекурсивним, тобто $x[n] = \sum_{k=1}^p a_k x[n-k]$.

В разі, коли фільтр є нерекурсивним, ми отримуємо модель *осереднення зі зсувом*: $x[n] = \sum_{k=0}^q b_k u[n-k]$.

Коли ми розглядаємо фільтр загального вигляду, з рекурсивною та нерекурсивною ланками, то маємо модель *авторегресії-осереднення зі зсувом*:

$$x[n] = \sum_{k=0}^q b_k u[n-k] + \sum_{k=1}^p a_k x[n-k].$$

Етапи оцінювання СГП випадкового сигналу авторегресійним методом:

1. Вибір параметричної моделі часового ряду, яка відповідає припущенням про аналізований сигнал.
2. Ідентифікація (оцінка) параметрів моделі – коефіцієнтів фільтра.
3. По оціненим значенням параметрів розраховують СГП. Маючи коефіцієнти фільтра, можна пов'язати z -перетворення вхідного шуму та вихідного сигналу через характеристичну функцію фільтра, розрахувати кореляційну функцію вихідного сигналу, та за теоремою Вінера-Хінчіна оцінити спектр.

6.7. Підсумки до розділу 6 та рекомендована література

В цьому розділі були коротко розглянуті основи аналізу випадкових сигналів. Основною відмінністю від аналізу детермінованих сигналів є те, що в загальному випадку випадковий сигнал може бути адекватно описаний та проаналізований лише на основі всього ансамблю реалізацій. На практиці частіше за все приймається гіпотеза про те, що аналізований сигнал є ергодичним та стаціонарним, і тому для отримання оцінок його параметрів достатньо провести лише один дослід та зареєструвати лише одну реалізацію.

Були отримані вирази для розрахунку основних ймовірнісних характеристик неперервних та дискретних стаціонарних ергодичних випадкових сигналів, таких як середнє значення, дисперсія та автокореляційна функція. Також були розглянуті параметричні та непараметричні методи оцінки спектральної густини потужності випадкових сигналів, основним з яких є періодограмний метод Уелча та авторегресійні методи.

Варто зауважити, що переважна більшість сигналів, які зустрічаються в реальному житті, є сигналами, що породжуються дуже складними технічними або біологічними системами. Вони є результатом проходження багатьох процесів, які складно описати точно, тому природнім є розгляд результуючих сигналів як випадкових і їх аналіз із використанням математичного апарату теорії ймовірностей. Але необхідно пам'ятати, що математичні методи аналізу детермінованих сигналів також є достатньо розвинутими та потужними, і що одну (відому досліднику) реалізацію випадкового сигналу можна вважати сигналом детермінованим. Отже, опанування математичним апаратом всіх доступних методів аналізу сигналів дає досліднику необхідну свободу у виборі та адаптації підходів до вирішення кожної конкретної задачі.

Теорія аналізу випадкових процесів – дуже широка та розвинута галузь аналізу. Для більш глибокого опанування нею можна рекомендувати ґрунтовні та добре написані книги Андерсона [33] та Свешнікова [34].

6.8. Питання для самоперевірки

1. Що таке випадковий сигнал, випадкова послідовність, випадкова функція, і як вони пов'язані між собою?
2. Що таке ансамбль реалізацій випадкового сигналу, як його отримати та коли для опису сигналу достатньо мати лише одну реалізацію?
3. Які існують основні ймовірнісні характеристики випадкових сигналів?
4. За якими ознаками можна класифікувати випадкові сигнали?
5. Що таке зміщеність та змістовність оцінки?
6. Чому бажано аналізувати стаціонарні та ергодичні випадкові сигнали та які характеристики таких сигналів можна отримати?
7. Які методи існують для оцінки спектральної густини потужності випадкового сигналу? В чому відмінність параметричних та непараметричних методів оцінки?

7. ОСНОВИ АНАЛІЗУ БАГАТОВИМІРНИХ СИГНАЛІВ

Equation Chapter 1 Section 7

7.1. Зображення як багатовимірні сигнали

В попередніх розділах мова йшла про сигнали, які були функціями одної змінної, як правило, часу $s(t)$. Такі сигнали називаються одновимірними. До цього класу сигналів відносяться всі сигнали, які вимірюються з використанням одного вимірювального каналу. Багатовимірними називаються сигнали, значення яких залежить не від одної змінної, а від кількох. Найбільш розповсюдженим прикладом багатовимірного сигналу є зображення. В такому сигналі значення сигналу є функцією двох координат в просторі $I_1(x, y)$. Наприклад, якщо мова іде про сірошкальне зображення, то значення сигналу є яскравістю точки в просторі. Як правило, яскравість – число від 0 до 1, де 0 відповідає точці чорного кольору (відсутність яскравості), а 1 – відповідає точці білого кольору (максимальна яскравість). Проміжні значення між 0 та 1 говорять про насиченість кольору: 0.5 відповідає сірому кольору, 0.2 – темно-сірому, 0.8 – світло-сірому і т.д.

Всі методи аналізу та обробки одновимірних сигналів, які вивчалися до цього, можуть бути застосовані до багатовимірних сигналів, але з урахуванням залежності значення сигналів від більше ніж одної змінної.

Отримання, обробка та аналіз зображень розповсюджені в багатьох галузях науки і техніки, і набувають все більш широкого розвитку.

Застосування аналізу зображень:

- аналіз та обробка двовимірних зображень (фотографій) при стисненні, кодуванні, розпізнаванні об'єктів, покращенні якості зображень;
- аналіз та обробка рухомих зображень (відеопослідовностей) при стисненні відео, автоматичному контролі якості продукції, стеженні за транспортними засобами;
- біомедичні застосування (аналіз та розпізнавання зображень, які отримані в результаті УЗД, МРТ, КТ та ін. технік медичної інтроскопії).

Основні напрями досліджень з обробки та аналізу зображень:

- вивчення сприйняття зображень людиною та технічними системами;
- розробка та удосконалення техніки реєстрації зображень;
- розробка методів покращення якості зображень (підвищення контрастності, чіткості, позбавлення від шумів);
- розробка ефективних та швидких методів стиснення зображень;
- сегментація зображень (розділення зображення на складові частини);
- морфологічний аналіз та розпізнавання об'єктів на зображеннях.

Основними методами отримання зображень є такі:

- отримання зображень у видимому діапазоні довжин хвиль (фотографія, відеозйомка);

– термографія – в результаті вимірювання температурного поля отримують інфрачервоні зображення, які показують розподіл тепла по поверхні об'єкта. Зокрема, в медицині такі зображення використовуються при діагностуванні запалень, захворювань судин та онкологічних захворювань шкіри. В електроніці методом термографії можна визначити області локального розігріву плати, мікросхеми або електронного чіпа;

– ультразвукова інтроскопія – при цьому система п'єзовипромінювачів генерує і передає в об'єкт ультразвукові імпульси. Ці звукові хвилі проходять скрізь об'єкт і на границях між областями із різним акустичним опором відбувається часткове віддзеркалення (частина хвилі повертається в бік зонду), частина затухає, а частина проходить далі, поки не досягнуть іншої границі і т.д. Віддзеркалені хвилі реєструються приймачем та передаються в комп'ютер, де на основі часу приходу кожного віддзеркаленого імпульсу та швидкості розповсюдження звуку в об'єкті розраховують відстань від зонду до границі розподілу. Обчислені відстані та інтенсивності будуються у вигляді зображення на дисплеї. Часто такі зображення будуються в реальному часі та, наприклад, дозволяють спостерігати за роботою серця чи активністю плоду;

– рентгенівська комп'ютерна томографія – при цьому дослідженні набір детекторів випромінювання та джерело рентгенівських променів розташовуються по кільцю, всередину якого поміщають об'єкт (наприклад, пацієнта). Кільце обертається навколо пацієнта, і випромінювання, яке пройшло через пацієнта, реєструється детекторами, що розташовуються на протилежному боці кільця. За допомогою математичного перетворення Радона можна побудувати зображення перерізу тіла пацієнта в площині кільця використовуючи інтенсивності випромінювання, яке пройшло через пацієнта. Цей процес відбувається неперервно при повороті кільця та при зміщенні кільця вздовж пацієнта (по спіралі). При зміщенні пацієнта вздовж вісі кільця отримують багато перерізів, які у сукупності утворюють тривимірне зображення;

– ядерно-магнітний резонанс або магнітно-резонансна томографія (ЯМР або МРТ) – при цьому дослідженні пацієнта поміщають в сильне постійне магнітне поле. Через його тіло пропускають радіохвилі у вигляді коротких імпульсів, на які реагують спіни електронів в атомах водню. У відповідь на радіоімпульси атоми водню в тканинах пацієнта випромінюють радіосигнали, сила та місце походження яких визначаються комп'ютерною системою обробки. В результаті будується тривимірне зображення тіла пацієнта;

– гамма-ізотопна візуалізація (однофотонна емісійна комп'ютерна томографія, ОФЕКТ) – при цьому пацієнту вводять радіоактивний ізотоп, розпад якого супроводжується гамма-випромінюванням. Це випромінювання реєструється детекторами, підсилюється та використовується для побудови тривимірних карт та будується зображення розподілу ізотопу в тілі. В основному цей метод використовують для знаходження ділянок патологій кісток, при інфекційних та онкологічних захворюваннях.

– позитронно-емісійна томографія – при ній пацієнту вводять радіоактивний ізотоп, розпад якого супроводжується позитронним випромінюванням. При зустрічі позитрона з електроном вони анігілюють з утворенням двох гамма-квантів. Результуюче гамма-випромінювання реєструється детекторами та формується томографічне зображення.

7.2. Представлення неперервних та дискретних зображень

З точки зору теорії сигналів, *зображення* – це функція двох змінних $I_1(x, y)$, де x та y – координати на площині.

Значення $I_1(x, y)$ в довільній точці площини називається *яскравістю* (інтенсивністю) або *рівнем сірого* для зображення. Як і у випадку одновимірних сигналів, $I_1(x, y)$ називається неперервним зображенням в разі, коли координати x та y можуть приймати довільні значення, а дискретним – в разі, коли x та y визначені лише для деякої множини значень. Можна провести дискретизацію зображень, якщо ввести поняття частоти дискретизації F_s або відповідного кроку дискретизації T_s . В загальному випадку, частота дискретизації може бути різною по різних координатам (x, y) : F_{sx} , F_{sy} (з відповідними кроками дискретизації T_{sx} , T_{sy}). Як правило, дискретизація зображень проводиться не в часовій області, а в просторовій, і частота дискретизації показує, скільки відліків зображення буде виміряне на одиницю довжини по кожній координаті. Розмірність частоти дискретизації в цьому випадку буде $\frac{1}{m}$, а крок дискретизації буде виражений в одиницях довжини.

Можна ввести двовимірний аналог одновимірної дельта-функції Дірака $\delta(t)$:

$$\delta(x, y) = \begin{cases} +\infty, & x = 0, y = 0, \\ 0, & x \neq 0, y \neq 0. \end{cases}$$

З її використанням дискретне зображення, яке буде отримано при дискретизації неперервного зображення $I(x, y)$ з кроками T_{sx} , T_{sy} можна отримати так:

$$I_1(x_n, y_m) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} I_1(x, y) \delta(x - nT_{sx}, y - mT_{sy}).$$

Тут (x_n, y_m) – координати точки на зображенні, яка має інтенсивність $I_1(x_n, y_m)$.

Вводять також дискретний аналог функції Дірака для зображень:

$$\delta[k, l] = \begin{cases} 1, & k = 0, l = 0, \\ 0, & k \neq 0, l \neq 0. \end{cases}$$

З використанням цього виразу дискретизоване зображення можна записати так:

$$I_1[n, m] = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} I_1(x, y) \delta(k - n, l - m), \quad k, l \in Z.$$

Дискретне зображення задане своїми значеннями $I_1[n, m]$ в деяких точках площини. В реальних застосуваннях дискретне зображення містить скінченну кількість елементів, які називаються пікселями (pixel, від англійського picture element – елемент зображення). В цьому випадку дискретна функція $I_1[n, m]$ має лише скінченну кількість значень:

$$I_1[n, m], \quad n = \overline{0, N-1}, \quad m = \overline{0, M-1}.$$

Таке зображення можна представити у вигляді матриці, кожний елемент якої дорівнює значенню яскравості відповідного пікселя:

$$I_1[n, m] = \begin{bmatrix} I_1[0, 0] & I_1[0, 1] & \dots & I_1[0, M-1] \\ I_1[1, 0] & I_1[1, 1] & \dots & I_1[1, M-1] \\ \vdots & & \ddots & \\ I_1[N-1, 0] & I_1[N-1, 1] & \dots & I_1[N-1, M-1] \end{bmatrix}.$$

Нумерація пікселів подвійна, перший індекс відповідає номеру рядка, другий – номеру стовпця. Лівий верхній піксель зображення має номер $[0, 0]$. Всі методи аналізу та обробки зображень зводяться до математичних операцій над неперервною функцією $f(x, y)$, яка описує неперервне зображення, або над пікселями $I_1[n, m]$ (у випадку дискретного зображення).

Якщо значення яскравості кожного пікселя $I_1[n, m]$ може приймати лише значення 0 або 1, то таке зображення називається бінарним, і пікселі можуть бути тільки двох кольорів (як правило, чорного та білого). У випадку, коли значення $I_1[n, m]$ може бути дійсним числом від 0 до 1, то говорять, що таке зображення є сірошкальним (grayscale image), або зображенням в градаціях сірого.

Далі в цьому розділі будуть коротко розглянуті основні підходи до обробки та аналізу дискретних зображень.

7.3. Обробка зображень в просторовій області

Термін «просторова область» відноситься до площини, в якій задане зображення. При просторовій обробці всі методи маніпулюють безпосередньо величинами яскравостей пікселів зображення. Їх можна описати загальним виразом

$$I_2[n, m] = T(I_1[n, m]),$$

де T – деяке правило, яке ставить у відповідність кожному пікселю початкового зображення $I_1[n, m]$ відповідний піксель результуючого зображення

$I_2[n, m]$, який отриманий в результаті обробки. Для отримання $I_2[n, m]$ може використовуватися не тільки значення пікселя $I_1[n, m]$, але також і інші пікселі в деякому околі $I_1[n \pm \Delta_n, m \pm \Delta_m]$. Як правило, це прямокутний або квадратний окіл.

При обробці зображення центр цього околу зміщують від пікселя до пікселя і проводять розрахунки для кожного центрального пікселя.

Градаційні перетворення зображень

Якщо для того, щоб отримати яскравість пікселя обробленого зображення, використовується яскравість лише одного пікселя початкового зображення, то говорять про градаційні перетворення. Іншими словами, обробка проводиться в околі розміром 1×1 піксель. Якщо яскравості пікселів початкового та результуючого зображення рівні відповідно $s_{nm} = I_1[n, m]$, та $r_{nm} = I_2[n, m]$, то можна записати, що

$$r_{nm} = T(s_{nm}).$$

Тут T описує залежність, яка пов'язує яскравості кожного пікселя початкового та результуючого зображення.

Нехай ми маємо сірошкальне зображення, яскравості пікселів якого лежать в межах від 0 до 1. На рисунку 7.1 наведені приклади деяких залежностей T . Перетворення яскравості згідно з кривою на рис. 7.1а не приведе до зміни яскравості пікселя, оскільки крива T переводить значення яскравості s_{nm} в таке саме значення r_{nm} . Якщо змінювати яскравість за кривою на рис. 7.1б, то зміни будуть помітні. Видно, що яскравості s_{nm} в околі нуля (дуже темні пікселі) будуть перетворені в пікселі r_{nm} нульової яскравості. Так само, дуже світлі пікселі (в з яскравостями близькими до 1) стануть білими (будуть мати яскравість 1). Достатньо вузький діапазон яскравостей початкового зображення s_{nm} від 0.4 до 0.6 (сірі пікселі) перетвориться на дуже широкий діапазон яскравостей в результуючому зображенні від 0.1 до 0.9, тобто розтягнеться. Це буде приводити до того, що деталі, які були на початковому зображенні сірими та слабо розрізнялися по кольору, будуть більш чітко видимі на зображенні після обробки.

При перетворенні за залежністю з рисунку 7.1в всі пікселі з яскравістю від 0 до 0.5 будуть мати яскравість 0, а пікселі з яскравістю більше 0.5 матимуть яскравість 1. Отже, зображення з сірошкального стане чорно-білим: пікселі темніші за сірий колір стануть чорними, а світліше за сірий – білими.

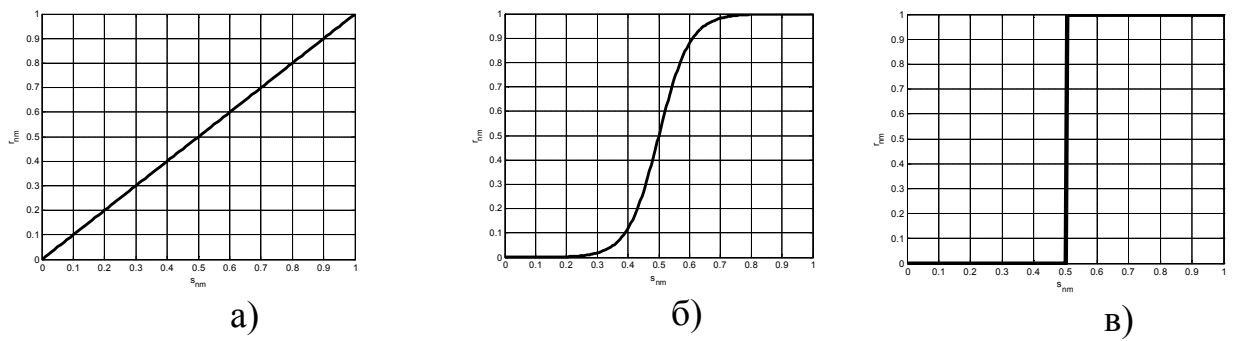


Рисунок 7.1 – Приклади залежностей T між яскравостями пікселів при градаційній обробці (пояснення в тексті)

Широковживаним видом градаційних перетворень є степеневі перетворення. При цьому яскравості пікселів результуючого та початкового зображення пов'язані за таким законом:

$$r_{nm} = c(s_{nm} + \varepsilon)^\gamma,$$

де c, ε, γ – деякі додатні константи.

Часто використовується більш проста залежність степеневого перетворення:

$$r_{nm} = c s_{nm}^\gamma.$$

Графіки для залежностей при різних показниках γ подані на рисунку 7.2.

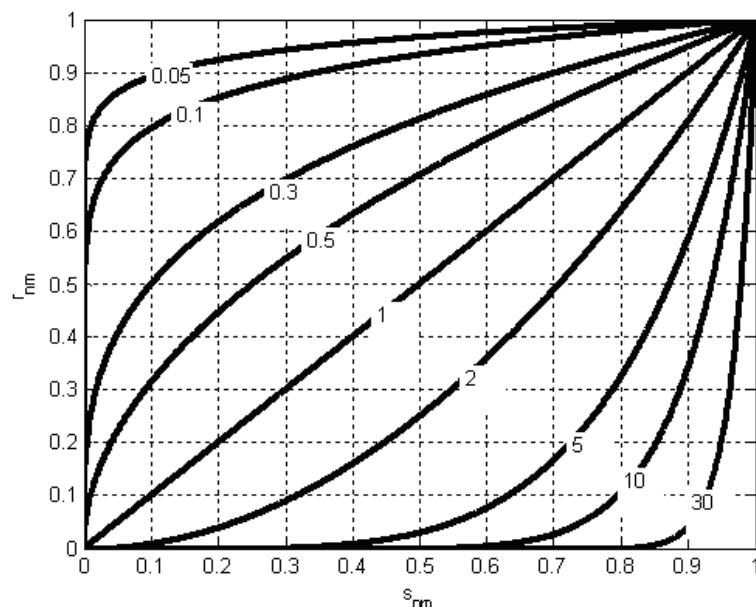


Рисунок 7.2 – Залежності між яскравостями пікселів вхідного та вихідного зображення при гамма-корекції для різних значень γ

Видно, що значеннях γ менше за одиницю, криві таких степеневих залежностей відображають вузький діапазон малих значень яскравостей пікселів вхідного зображення у широкий діапазон яскравостей пікселів результуючого зображення. Якщо значення показника степеню γ більше за одиницю, то відбувається протилежний ефект: вузький діапазон великих яскравостей відображається у широкий діапазон яскравостей пікселів результуючого зображення. Ця процедура називається *гамма-корекцією*.

Еквалізація гістограми

Гістограмою дискретного зображення називається дискретна функція $H(b_k) = \frac{N_k}{N}$, де b_k є k -им рівнем яскравості пікселя ($k = 0 \dots K-1$), N_k – кількість пікселів, які мають яскравість b_k , а N – кількість пікселів у всьому зображенні. Значення $H(b_k)$ є оцінкою імовірності появи пікселя яскравості b_k в зображенні.

Еквалізацію (лінеаризацію) гістограми проводять в тому випадку, коли в зображенні є багато пікселів зі схожими яскравостями, і мало пікселів з іншими яскравостями. На гістограмі ми будемо бачити, що на деяких проміжках яскравостей згруповано багато пікселів, в той час як деякі проміжки яскравостей майже не зайняті. При цьому деталі зображення, які зображені цими кольорами, складно розрізнити. Натомість існують такі проміжки яскравості, пікселів з якими взагалі немає на зображенні. Ці вільні проміжки яскравості можна «зайняти» для покращення якості зображення. Для цього роблять еквалізацію гістограм.

Якщо маємо піксель початкового зображення з яскравістю b_k , яка є k -им рівнем яскравості на гістограмі ($k = 0 \dots K-1$) то яскравість відповідного пікселя результуючого зображення буде розраховуватися

$$r_k = \sum_{p=0}^k H(b_p) = \sum_{p=0}^k \frac{N_p}{N}.$$

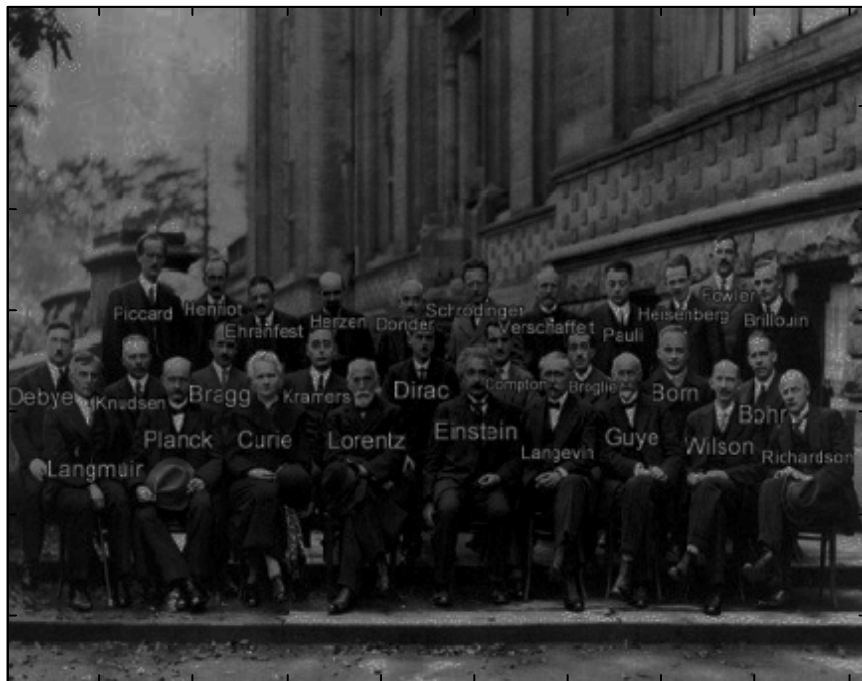
В результаті еквалізації гістограми яскравості пікселів на ній будуть розподілені рівномірно по всій шкалі яскравостей.

Наприклад, на рис. 7.3а наведене зображення, яке виглядає дуже темним. Дрібні деталі предметів та людей на ньому розрізнити складно, оскільки вони зображені схожими темними кольорами, які мало відрізняються один від одного. Гістограма цього зображення наведена на рис. 7.3б. На ній видно, що багато пікселів знаходяться в лівій частині шкали кольорів, що відповідає темним кольорам. Водночас, права частина шкали майже не зайнята, тобто світлих пікселів на зображенні немає. Цей вільний проміжок гістограми можна використати, щоб перенести туди яскравості деяких пікселів. Якщо гістограму цього зображення «розтягнути» на весь доступний діапазон яскравостей, то пікселі, які раніше мали

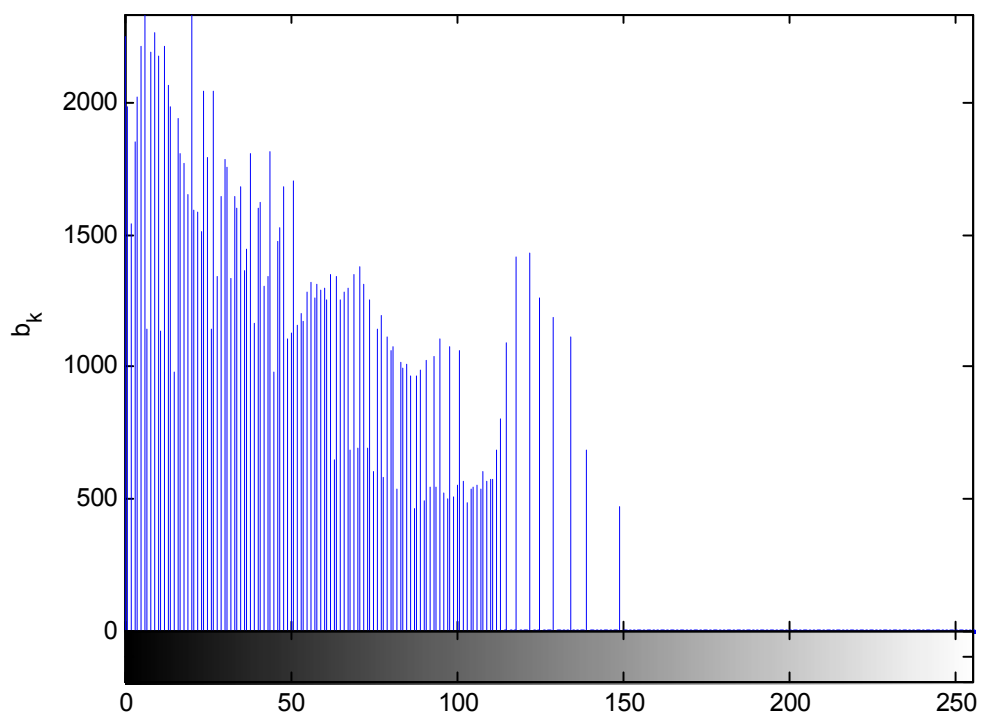
дуже схожі кольори (їх яскравості знаходились близько на шкалі яскравостей), будуть віддалені один від одного на більшу величину яскравості.

Якщо подивитись на зображення, видно, що діапазон яскравостей пікселів, які присутні на зображенні, розширився: на зображенні тепер є і темні, і світлі пікселі (рис. 7.4а). Тепер стало легше розрізнити деталі зображення, оскільки вони зображені більш контрастно. На гістограмі видно, що з зображенні присутні пікселі всіх яскравостей, і весь діапазон яскравостей тепер зайнятий.

Перевагою еквалізації гістограм є те, що цей метод легко автоматизується і не вимагає задавання ніяких додаткових параметрів для отримання покращеного зображення. Розрахунки для еквалізації гістограм такою достатньо нескладні.

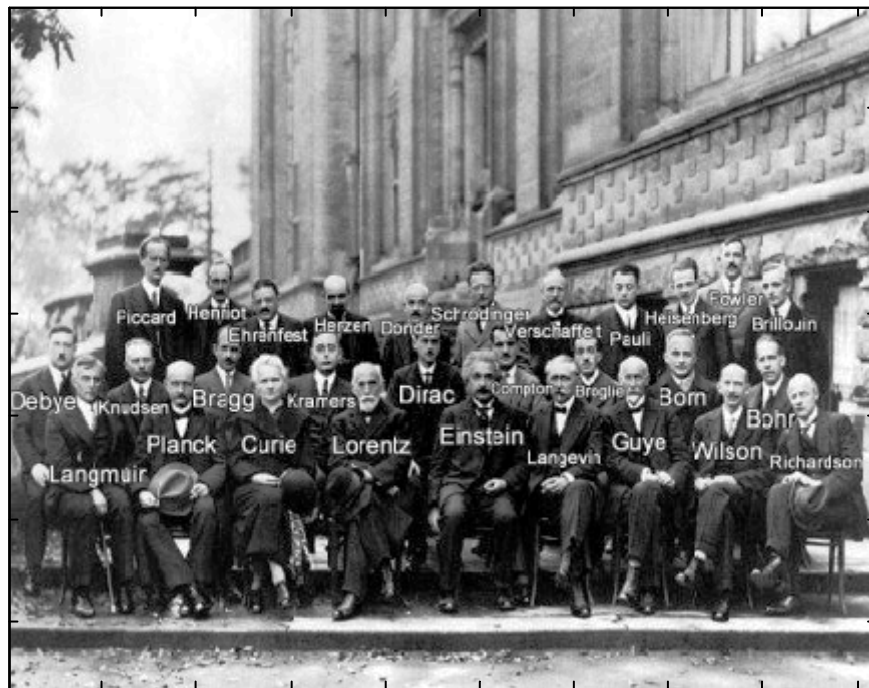


a)

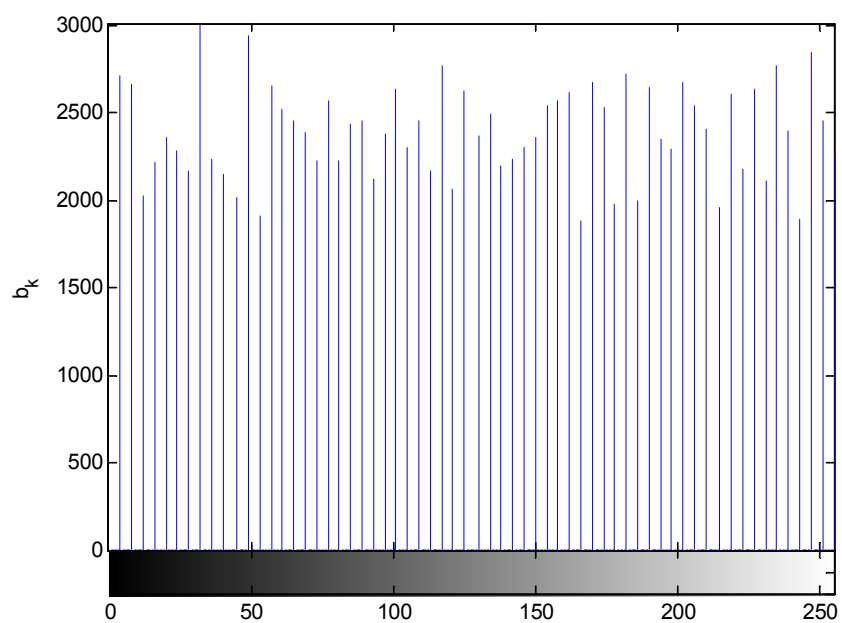


б)

Рисунок 7.3 – Початкове сірошкальне зображення (а) та його гістограма (б). градації сірого кольору кодуються числами від 0 до 255, оскільки для запису кольору пікселя використовується 8 біт



a)



б)

Рисунок 7.4 – Сірошкальне зображення (а) та його гістограма (б) після еквалізації

Просторова фільтрація

Просторова фільтрація зображень – це різновид обробки зображень в просторовій області, на відміну від класичного поняття «фільтрації», яка виконується із залученням спектральних характеристик фільтрів та спектрів зображень. При просторовій фільтрації для розрахунку яскравості пікселя результуючого зображення використовуються яскравості пікселів в деякому околі початкового зображення.

Схематично процес просторової фільтрації проілюстрований на рис. 7.5. Обробка виконується послідовно для кожного пікселя зображення. В початковому зображенні обирається окіл розміром 3×3 пікселя з центром в деякому пікселі $I_1[n, m]$. На основі значень яскравості пікселя $I_1[n, m]$ та його околу буде розрахована яскравість одного пікселя результуючого зображення $I_2[n, m]$. Для цього обирають «маску» коефіцієнтів, яка містить числа, на які мають смисл коефіцієнтів, з якими яскравість кожного пікселя з околу на початковому зображенні буде використана для отримання яскравості пікселя в результуючому зображенні. Розрахунок яскравості ведеться за формулою:

$$\begin{aligned} I_2[n, m] = & c_{-1, -1} I_1[n - 1, m - 1] + c_{-1, 0} I_1[n - 1, m] + c_{-1, 1} I_1[n - 1, m + 1] + \\ & + c_{0, -1} I_1[n, m - 1] + c_{0, 0} I_1[n, m] + c_{0, 1} I_1[n, m + 1] + \\ & + c_{1, -1} I_1[n + 1, m - 1] + c_{1, 0} I_1[n + 1, m] + c_{1, 1} I_1[n + 1, m + 1]. \end{aligned}$$

Видно, що маска центрується в пікселі з номером $[n, m]$, і яскравість цього пікселя множиться на коефіцієнт $c_{0, 0}$, що знаходиться в центрі маски.

При використанні маски розміром $A \times B$ (A і B повинні бути непарними) загальний вираз для отримання яскравості кожного пікселя результуючого зображення має вигляд:

$$\begin{aligned} I_2[n, m] = & \frac{1}{C} \sum_{p=-A}^A \sum_{s=-B}^B c_{p, s} I_1[n + p, m + s] = \\ & \frac{\sum_{p=-A}^A \sum_{s=-B}^B c_{p, s} I_1[n + p, m + s]}{\sum_{p=-A}^A \sum_{s=-B}^B c_{p, s}}. \end{aligned} \quad (7.1).$$

Константа C в попередньому виразі є нормувальним множником. Вона дорівнює сумі всіх коефіцієнтів маски, тому при діленні на неї сума коефіцієнтів при всіх яскравостях пікселів дорівнює одиниці. Це приводить до того, що яскравість пікселя відфільтрованого зображення буде не більше, ніж максимальна

припустима для даного зображення яскравість. Нормувальний множник застосовується в тому випадку, якщо сума коефіцієнтів маски не дорівнює нулю.

При просторовій фільтрації з допомогою масок, вираз (7.1) повинен бути застосований по черзі для кожного пікселя початкового зображення. Треба мати на увазі, що при проходженні маски по краю зображення деякі пікселі можуть бути відсутні. Тому розрахунок результуючих пікселів повинен виконуватися за особливою процедурою для уникнення крайових ефектів. При цьому зображення або дзеркально продовжується, або добудовується деякий окіл по краю зображення (наприклад, він заповнюється пікселями з нульовими або середніми яскравостями). Також можна не проводити розрахунок яскравостей для таких пікселів, але в цьому випадку розмір отриманого після фільтрації зображення буде меншим на два стовпця та два рядка.

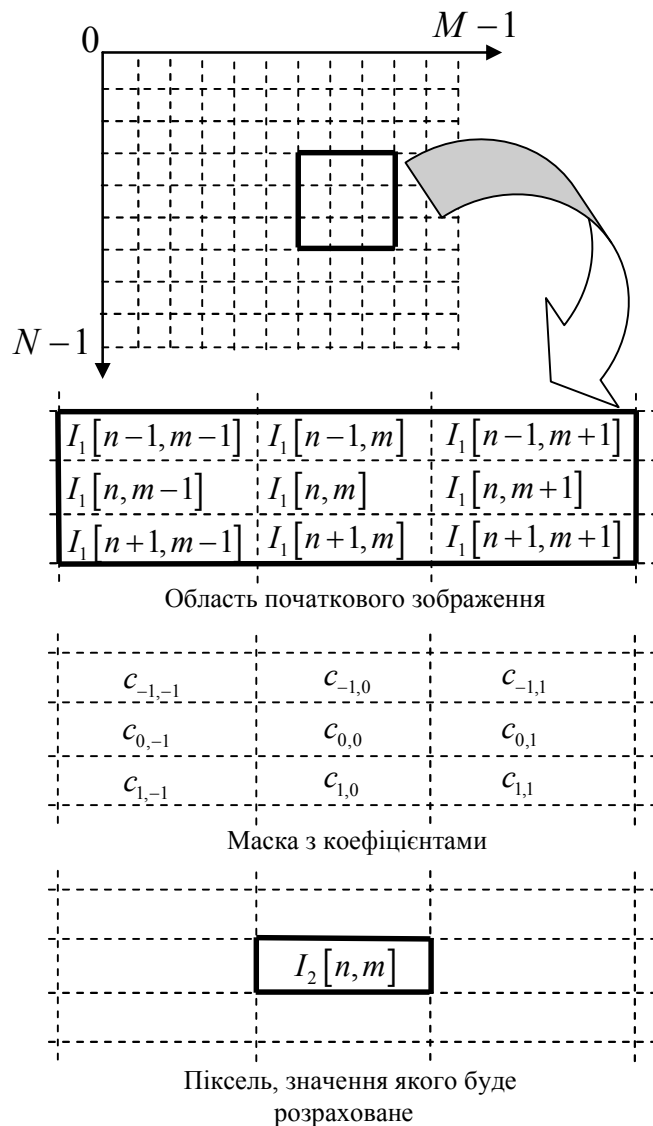


Рисунок 7.5 – Ілюстрація процесу просторової фільтрації

На рис. 7.6 наведені приклади масок, які часто використовуються на практиці. При використанні маски з рис. 7.5а виконується згладжування зображення, оскільки центральний піксель результуючого зображення буде мати яскравість, яка є середньою для околу відповідного пікселя на початковому зображенні. Маска на рис. 7.5б реалізує оператор Лапласа для зображення (взяття другої часткової похідної по двом напрямкам). Її застосування підкреслює розриви рівнів яскравостей на зображенні, і в результаті підвищується різкість наявних на зображенні границь. Послідовне використання двох масок з рис. 7.5в та 7.5г еквівалентне розрахунку градієнта зображення; застосування цих масок ще називають оператором Собеля. Розрахунок градієнтів використовують для покращення видимості контурів на зображенні.

$$\begin{array}{cc}
 \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \\
 \text{а)} & \text{б)} \\
 \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} & \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \\
 \text{в)} & \text{г)}
 \end{array}$$

Рисунок 7.6 – Приклади масок: а) згладжувальна; б) підвищення різкості; в) і г) оператор Собеля

Медіанна фільтрація

При медіанній фільтрації для отримання яскравості пікселя результуючого зображення використовується окіл відповідного пікселя початкового зображення. Із значень яскравостей пікселів околу отримують послідовність, в якій вони розташовані від найменшої до найбільшої. Для цієї послідовності знаходиться медіана, тобто визначається, який піксель знаходиться тому місці в послідовності, яке відповідає половині її довжини. Наприклад, якщо в околі 9 пікселів, то медіаною буде той піксель, який іде п'ятим у ранжованому ряді. Яскравість цього пікселя буде значенням яскравості пікселя відфільтрованого зображення.

7.4. Підхід до обробки зображень в частотній області

Аналогічно до того, як для одновимірного сигналу в часовій області можна знайти його спектральне представлення за Фур'є, можна знайти спектр зображення.

Якщо маємо неперервне зображення $I_1(x, y)$, то його спектр за Фур'є буде визначатися:

$$F(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} I_1(x, y) e^{-j(ux+vy)} dx dy, \quad (7.2)$$

де u, v – просторові частоти, $\frac{1}{M}$.

Обернене перетворення Фур'є для зображення:

$$I_1(x, y) = \frac{1}{4\pi^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(u, v) e^{j(ux+vy)} du dv. \quad (7.3)$$

Якщо зображення дискретне розмірністю $N \times M$, то аналоги формул (7.2)-(7.3) будуть мати вигляд:

$$F[k, p] = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} I_1[n, m] e^{-j2\pi \left(\frac{kn}{N} + \frac{pm}{M} \right)}, \quad (7.4)$$

$$I_1[n, m] = \sum_{k=0}^{N-1} \sum_{p=0}^{M-1} I_1[n, m] e^{j2\pi \left(\frac{kn}{N} + \frac{pm}{M} \right)}. \quad (7.5)$$

Тут k, p – номери гармонічних спектральних складових зображення, $k = \overline{0, N-1}$, $p = \overline{0, M-1}$.

Як і для одновимірного випадку, спектр дискретного зображення є періодичною функцією частот k, p з періодами, які дорівнюють розмірностям зображення: $F[k \pm qN, p \pm cM] = F[k, p]$ для цілих q, c .

Аналогічно до одновимірного випадку, для комплексного спектру зображення можна отримати амплітудний та фазовий спектри:

$$|F(u, v)| = \sqrt{\operatorname{Re}(F(u, v))^2 + \operatorname{Im}(F(u, v))^2},$$

$$\phi(u, v) = \arg(F(u, v)) = \operatorname{arctg} \frac{\operatorname{Im}(F(u, v))}{\operatorname{Re}(F(u, v))}.$$

Фільтрація зображень в частотній області полягає у наступному. В двовимірній частотній області задається комплексна частотна характеристика фільтра $H(u, v)$. Як і для одновимірного випадку, можна створити фільтри нижніх (ФНЧ), верхніх (ФВЧ) частот, смугові (СФ) та загороджувальні (ЗФ) фільтри. Фільтрація зображення еквівалентна проходженню його через відповідний лінійний фільтр. В частотній області для того, щоб отримати спектр результуючого зображення $H(u, v)$, необхідно перемножити спектр початкового зображення $F_1(u, v)$ на КЧХ фільтра:

$$F_2(u, v) = H(u, v) \cdot F_1(u, v).$$

Для отримання відфільтрованого зображення необхідно виконати обернене перетворення Фур'є.

7.5. Підсумки до розділу 7 та рекомендована література

В цьому розділі стисло подані основи аналізу двовимірних сигналів – зображень. Зображення було представлено як функція, яка залежить від двох просторових змінних, і яка аналогічно до одновимірного випадку може бути неперервною і дискретною. Розглянуто два способи обробки зображень: в просторовій та частотній областях.

Просторова обробка – коли для отримання пікселів результуючого зображення виконуються математичні операції з одним або кількома пікселями початкового зображення; це зміна яскравостей пікселів, еквалізація гістограм або застосування масок. Обробка в частотній області – фільтрація зображення за допомогою фільтрів різних типів, яка приводить до зміни характеристик спектру зображень. Підбираючи типи фільтрів, якими будемо користуватись, та виконуючи апроксимацію їх частотних характеристик, після фільтрації можна досягти зменшення рівня шумів. В розділі подані вирази для гістограмної та градаційної обробки зображень, а також вирази для прямого та оберненого перетворення Фур'є дискретного та неперервного зображення.

До тематик, які потребують значно більшого обсягу матеріалу і більш докладного вивчення, відносяться: обробка кольорових зображень, вейвлет-аналіз зображень, стиснення зображень, морфологічна обробка зображень, сегментація зображень. Також багато методів застосовується для обробки рухомих зображень (відео).

З початковими відомостями по цим тематикам можна ознайомитись в фундаментальній праці Р. Гонсалеса та Р. Вудса «Цифрова обробка зображень» [35]. З особливостями реалізації методів обробки та аналізу зображень в середовищі Матлаб можна ознайомитись в книзі Р.Гонсалес, Р.Вудс, С.Еддінс «Цифрова обробка зображень в середовищі Матлаб» [36].

7.6. Питання для самоперевірки

1. Які існують застосування обробки та аналізу зображень?
2. Як описуються неперервні та дискретні зображення?
3. В чому сутність градаційних перетворень зображень, як вони виконуються?
4. В чому сутність еквалізації гістограм зображень, як вона виконується?

5. В чому сутність просторової фільтрації, як вона виконується? Які можуть бути маски? Як можна врахувати крайові ефекти при роботі з маскою?

6. В чому сутність медіанної фільтрації, як вона виконується?

7. Записати вираз для прямого та оберненого перетворень Фур'є неперервних та дискретних зображень. Який фізичний зміст мають величини, які входять до цих виразів?

8. В чому сутність частотної фільтрації? Записати вираз для отримання спектру відфільтрованого зображення.

8. НЕЛІНІЙНИЙ ТА МУЛЬТИВАРІАТИВНИЙ АНАЛІЗ СИГНАЛІВ

Equation Chapter 1 Section 8

8.1. Загальна класифікація методів аналізу сигналів

До цього часу обговорювалися лише окремі методи аналізу сигналів, наприклад, спектральний аналіз, аналіз за Уолшем та ін. В цьому розділі буде зроблена спроба дати загальну характеристику методів аналізу та описати приклади методів, які є представниками різних підходів до аналізу сигналів.

За кількістю сигналів, які аналізуються одночасно, виділяють такі групи методів аналізу:

Унівариативний аналіз – аналізується один сигнал, який характеризує стан об'єкта.

Мультиваріативний аналіз – аналізується кілька (більше двох) сигналів, що у сукупності характеризують стан одного або кількох об'єктів.

Також окремо виділяють *бівариативний* аналіз, при якому аналізуються два сигнали, які характеризують один або два різні об'єкти.

Зазвичай методами мульти- та бівариативного аналізу досліджуються одночасно два та більше каналів сигналу від одного об'єкта. Але останніми роками все більшого розповсюдження набуває застосування цих методів до дослідження взаємодії між двома об'єктами (системами або процесами). При цьому одночасно реєструються сигнали, які характеризують роботу двох систем, а на основі методів мультиваріативного аналізу визначають вплив одної системи на іншу, синхронність та зв'язність їх роботи та ін.

За тим, яким чином пов'язані характеристики сигналу та результат його перетворення тим чи іншим методом, можна виділити такі групи методів:

Лінійні методи аналізу – величина, що характеризує властивості сигналу, змінюється пропорційно зміні самого сигналу.

Нелінійні методи – виявляють відсутність лінійного зв'язку між сигналом та розрахованими за певним методом параметрами. Ці методи здатні знайти такі характеристики сигналу, які значно змінюються при незначній зміні величини самого сигналу. Ці методи також можуть виявити такі зміни в сигналі, які не відображаються в величині сигналу, а відбуваються, наприклад в взаємному розташуванні тих чи інших патернів сигналу, або у зв'язках між величиною сигналу в різні моменти часу. В своїй більшості нелінійні методи залучають представлення сигналу в фазовому просторі.

8.2. Уніваріативний аналіз

Лінійні методи

До лінійних методів уніваріативного аналізу можна віднести спектральний аналіз сигналів за Фур'є. Перетворення Фур'є – лінійне за визначенням, оскільки визначається інтегралом. При цьому перетворенні досліджується спектральний склад сигналу: визначаються амплітуди і початкові фази синусоїдальних коливань, які складають сигнал.

Певним розширенням можливостей аналізу за Фур'є є вейвлет-аналіз. При розкладі сигналу за вейвлетами відбувається не виявлення в сигналі синусоїдальних компонент, а масштабно-часовий аналіз хвильової структури сигналу. При цьому хвильові коливання, на які ведеться розклад сигналу при вейвлет-аналізі, можуть мати різну форму.

Нелінійні методи

До цієї групи методів відносяться, в першу чергу, аналіз ентропії сигналу (ентропія перестановок, ентропія вибірок, наближена ентропія), хаотичності і непередбачуваності значень дискретного сигналу (фрактальні розмірності, експоненти Ляпунова), а також різні методи оцінки «складності» сигналу (складність за Лемпелем-Зівом, кодова довжина).

Аналіз ентропії перестановок

В якості прикладу нелінійного уніваріативного аналізу розглянемо аналіз *ентропії перестановок* (permutation entropy).

Ентропія перестановок — це характеристика хаотичності сигналу та невпорядкованості його значень, яка визначається на основі підходів символічної динаміки.

Ентропія перестановок може бути розрахована лише для дискретного сигналу. Її розрахунок ґрунтується на визначенні відносної кількості патернів в сигналі, які можуть бути утворені при всіх можливих перестановках кількох різних чисел. Перед розрахунком обирається порядок ентропії перестановок m (order) – це кількість відліків сигналу, які будуть входити в один патерн. Наприклад, на рис. 8.1 подані всі можливі патерни, які можна утворити при перестановках $m = 3$ різних за величиною чисел a, b, c . Всього існує $m!$ можливих перестановок. В практичних застосуваннях рекомендується брати порядок від третього до сьомого.

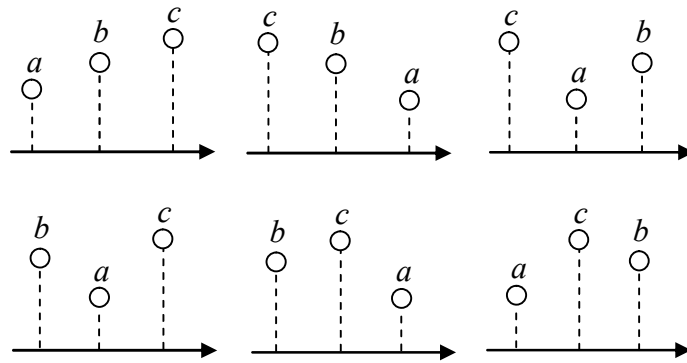


Рисунок 8.1 – Патерни розташування значень відліків сигналу, які будуть визначатися при розрахунку ентропії перестановок третього порядку

Після визначення порядку ентропії перестановок обирають величину затримки l (lag), яка буде визначати інтервал між відліками сигналу, які будуть формувати один патерн. Якщо затримка дорівнює 1, то відліки будуть братися підряд, якщо затримка дорівнює 2, то інтервал між відліками в патерні буде рівний двом крокам дискретизації, а відліки будуть братися через один, і т.д. (рис. 8.2)

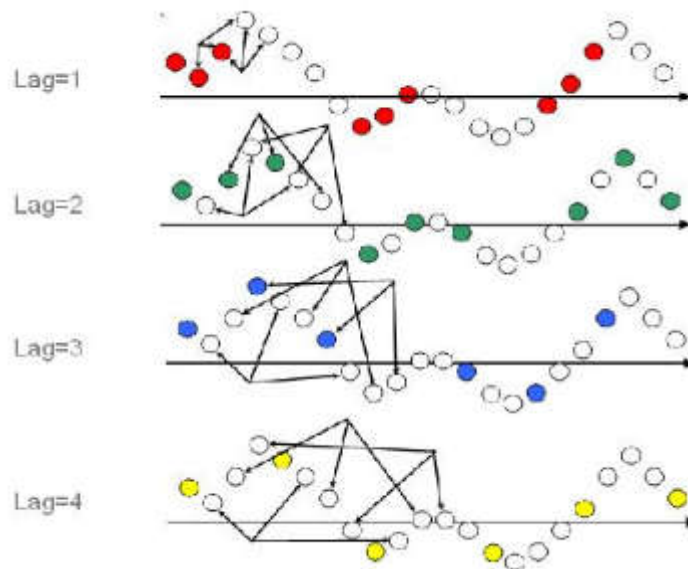


Рисунок 8.2 – Патерни в сигналі для розрахунку ентропії перестановок порядку $m = 3$ при різних значеннях затримки

Після того, як обрано порядок ентропії перестановок m та затримку l , в сигналі необхідно знайти відносну кількість кожного з $m!$ можливих патернів. Кожний патерн буде складатися з m відліків сигналу, які взяті з інтервалом l .

Ентропію перестановок розраховують за формулою:

$$PE(m, l) = - \sum_{i=1}^{m!} p(\pi_i) \log p(\pi_i),$$

де $p(\pi_i)$ – відносна кількість появи патерну π_i .

Теоретично можливе максимальне значення ентропії перестановок дорівнює $\log m!$, тому часто використовують нормовану величину

$$PE(m, l) = -\frac{1}{\log m!} \sum_{i=1}^{m!} p(\pi_i) \log p(\pi_i),$$

яка лежить в проміжку від 0 до 1. Величина $PE(m, l)$ близька до нуля відповідає сигналу, значення якого повністю передбачувані, наприклад, зростають чи спадають. Якщо для деякого сигналу отримана $PE(m, l)$ близька до одиниці, це означає, що сигнал містить всі можливі патерни з однаковою імовірністю, отже, є сигнал є випадковим.

Аналіз рекурентності

Ще одним нелінійним уніваріативним методом аналізу сигналів є аналіз рекурентності. При цьому вважається, що в сигналі існує деяка повторювальність чи циклічність поведінки, при якій ділянки сигналів, що знаходяться на деякій відстані в часі, можуть бути схожі між собою. Ця періодичність може бути виявлена в фазовому просторі, де стану системи в певний момент часу відповідає одна точка. При зміні стану системи ця точка рухається по деякій траєкторії в фазовому просторі, і ця траєкторія характеризує поведінку системи. При аналізі рекурентності вважається, що той сигнал, який підлягає аналізу, є відображенням динаміки деякої системи у фазовому просторі.

Спочатку визначається або задається розмірність фазового простору N , в якому буде відбуватися аналіз. Після цього із відліків сигналу $x[n] = [x[0], x[1], x[2], x[3], \dots, x[n], x[n+1], \dots]$ формуються «вкладені вектори» вигляду:

$$u_n = (x[n], x[n-1], x[n-2], \dots, x[n-(N-1)]).$$

Такі вектори є точками в N -вимірному фазовому просторі, які відповідають стану системи в кожний момент часу.

Наприклад, для випадку розмірності фазового простору $N=3$:

$$u_0 = (x[2], x[1], x[0])$$

$$u_1 = (x[3], x[2], x[1])$$

$$u_2 = (x[4], x[3], x[2])$$

$$u_3 = (x[5], x[4], x[3])$$

Маючи набір із M таких векторів, можна побудувати матрицю $R(i, j)$ з елементами, які розраховуються за таким правилом:

$$R(i, j) = \theta(\varepsilon - \|u_i - u_j\|), \quad i, j = 1 \dots M,$$

де $\theta(s)$ – функція Хевісайда,

$$\theta(s) = \begin{cases} 0, & s < 0 \\ 1, & s \geq 0, \end{cases}$$

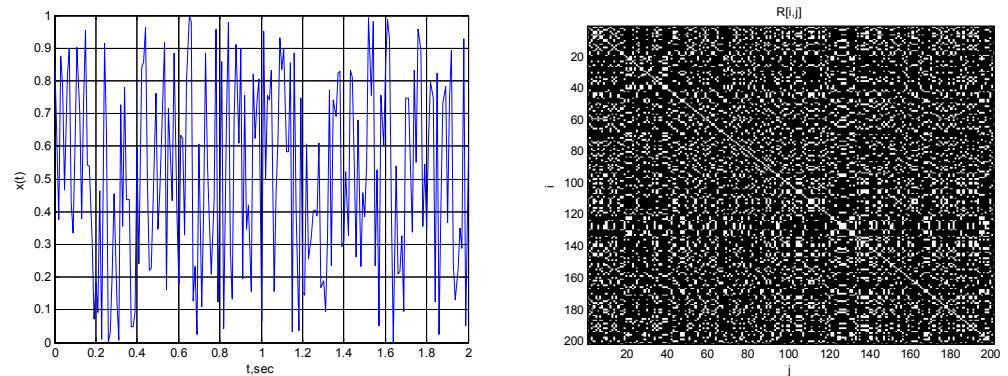
ε – деяке число.

В цій формулі для розрахунку $R(i, j)$ використовується порівняння відстані між двома векторами u_i та u_j в N -вимірному фазовому просторі з деякою граничною відстанню ε . Відстань між векторами визначається деякою нормою, як правило, евклідовою. Елементами матриці $R(i, j)$ будуть числа 0 або 1 залежно від того, чи знаходяться вектори u_i та u_j далі чи ближче за значення ε .

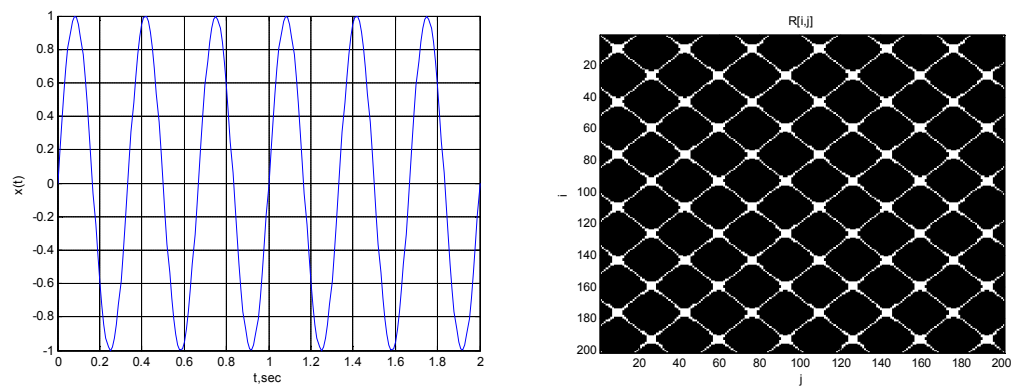
Вектори u_i та u_j відповідають станам системи в моменти часу з індексами i та j відповідно. Величина ε є радіусом N -вимірної гіперсфери в фазовому просторі з центром в точці u_i . Якщо система в деякий момент часу j знаходиться у стані, який схожий на стан в момент часу i , то вектор u_j буде знаходитись в просторі близько до вектора u_i . Якщо вектор u_j потрапить в окіл, що обмежений сферою з радіусом ε , то аргумент функції Хевісайда буде додатним, і $R(i, j)$ буде дорівнювати 1. Якщо стан системи в момент часу j буде не схожий на стан в момент i , то відстань між векторами u_i та u_j буде більша за ε , аргумент функції Хевісайда буде від’ємний, і значення $R(i, j)$ буде рівним 0.

Якщо зобразити елементи матриці $R(i, j)$ на площині, закодувавши значення «0» чорним кольором, а значення «1» – білим, то отримаємо «рекурентний графік». Він допомагає візуалізувати результат аналізу рекурентності та отримати загальну уяву щодо властивостей сигналу, який аналізується.

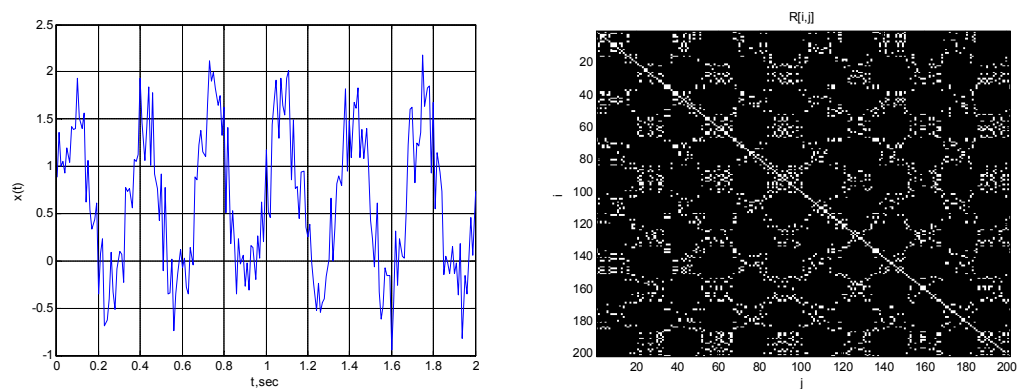
На рисунку 8.3 подано графіки сигналів та рекурентних графіків для них.



а)



б)



в)

Рисунок 8.3 – Сигнали та їх рекурентні графіки: а) випадковий сигнал, б) синусоїдальний сигнал, в) зашумлений синусоїдальний сигнал

Змінюючи значення ε , можна задавати поріг «близькості» векторів у просторі, і таким чином задавати, які стани системи будуть вважатися схожими при аналізі.

Вертикальні або горизонтальні лінії на рекурентному графіку говорять про те, що характер сигналу не змінювався протягом деякого часу. Діагональні лінії говорять про те, що різні ділянки сигналу змінюються схоже. Аналіз рекурентних графіків вимагає певних навичок і не є тривіальним. Часто використовують різні чисельні показники, які розраховані з використанням значень матриці $R(i, j)$.

Серед найбільш розповсюджених кількісних характеристик рекурентних графіків є такі:

- рівень рекурентності:

$$RR = \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M R(i, j).$$

- детермінізм:

$$D = \frac{\sum_{d=d_{\min}}^M d \cdot P(d)}{\sum_{i=1}^M \sum_{j=1}^M R(i, j)},$$

де $P(d)$ – функція густини розподілу значень довжин d діагональних ліній на рекурентному графіку;

- ламінарність:

$$L = \frac{\sum_{v=v_{\min}}^M v \cdot P(v)}{\sum_{i=1}^M \sum_{j=1}^M R(i, j)},$$

де $P(v)$ – функція густини розподілу значень довжин v вертикальних ліній на рекурентному графіку.

Аналіз флуктуацій, позбавлених тренду (Detrended Fluctuation Analysis)

Цей вид аналізу широко використовується для дослідження нестационарних сигналів на предмет наявності в них тривалих кореляцій (long-term correlations). Стандартна процедура аналізу флуктуацій, позбавлених тренду (Detrended Fluctuation Analysis, DFA) складається з таких кроків:

1. Від початкового сигналу $x[n]$ віднімається його середнє значення \bar{x} , і отриманий сигнал інтегрується:

$$y[k] = \sum_{n=1}^k (x[n] - \bar{x}).$$

2. Проінтегрований сигнал $y[n]$ розбивається на M сегментів однакової довжини N , які не перекриваються.

3. В кожному сегменті на основі методу найменших квадратів розраховується апроксимаційний поліном $\tilde{y}[n]$. Як правило, це поліном першого ступеня.

4. На кожному з M сегментів розраховують середній квадрат відхилення між сигналом $y[n]$ та апроксимаційним поліномом $\tilde{y}[n]$:

$$d_m = \frac{1}{N} \sum_{n=1}^N (y[n] - \tilde{y}[n])^2.$$

5. Середні квадрати відхилень осереднюються по всім сегментам:

$$F(M) = \sqrt{\frac{1}{M} \sum_{m=1}^M d_m}.$$

6. Кроки 2-5 повторюють для різних кількостей розбиття сигналу на сегменти, щоб отримати залежність середньоквадратичного відхилення F від M .

Для аналізу використовують тангенс кута нахилу лінійної частини графіка залежності $\log(F(M))$ від $\log(M)$ (log-log plot). Ця величина називається масштабною експонентою (scaling exponent) α . Залежно від її значення сигнал має такі властивості:

$\alpha < 0.5$ - анти-корельований сигнал;

$\alpha = 0.5$ - некорельований сигнал, білий шум;

$\alpha > 0.5$ - корельований сигнал;

$\alpha \approx 1$ - шум виду $\frac{1}{f}$, рожевий шум;

$\alpha > 1$ - нестационарний сигнал;

$\alpha \approx 1.5$ - Броунівський рух.

8.3. Біваріативний аналіз

Метою біваріативного аналізу є знаходження або оцінювання зв'язку між двома об'єктами (системами або процесами). При цьому апіорі вважається, що такий зв'язок існує (або його наявність перевіряють з використанням сурогатних даних). При біваріативному аналізі одночасно досліджують пару сигналів з метою визначити їх спільні характеристики. Часто ці два сигнали надходять від різних об'єктів, і при аналізі необхідно визначити, наскільки узгодженою, синхронною та взаємозалежною є активність цих двох систем.

Одним із простих методів біваріативного аналізу є дослідження взаємної кореляції між двома сигналами. Раніше було розглянуто дві характеристики кореляції, що чисельно описують, наскільки два сигнали змінюються разом порівняно з тим, наскільки вони змінюються окремо кожен. Перша – коефіцієнт кореляції, другий – взаємнокореляційна функція.

Ще один метод біваріативного аналізу – *кросс-спектральний аналіз* за Фур'є (його також називають *взаємносектральний аналіз*, або аналіз взаємної спектральної густини потужності). Цей аналіз дозволяє дослідити зв'язок між двома сигналами як функцію частоти. Передумовою для того, щоб вважати результати

цього аналізу змістовними, є припущення щодо того, що піки в спектрах обох сигналів на одних і тих самих частотах з'являються в результаті того, що процеси є пов'язаними один з одним.

Якщо маємо два сигнали $x(t)$ та $y(t)$ з крос-кореляційною функцією

$$R_{xy}(\tau) = \int_{-\infty}^{+\infty} x(t)y(t-\tau)dt,$$

то їх крос-спектр визначається за формулою:

$$F_{xy}(j\omega) = \int_{-\infty}^{+\infty} R(\tau)e^{-j\omega\tau}d\tau.$$

Ще одним методом біваріативного аналізу є дослідження *спектральної когерентності*. Вона визначається за формулою:

$$\gamma_{xy}(\omega) = \frac{F_{xy}(j\omega)}{\sqrt{F_{xx}(j\omega)F_{yy}(j\omega)}},$$

де $F_{xx}(j\omega)$ та $F_{yy}(j\omega)$ є спектральними густинами потужності сигналів $x(t)$ та $y(t)$ відповідно. Квадрат спектральної когерентності показує, наскільки суттєвою є потужність спектральних складових обох сигналів на однакових частотах.

Фазова синхронізація

Фазова синхронізація між двома сигналами є явищем встановлення та підтримання постійної різниці фаз між ними. З точки зору теорії, про поняття фази можна вести мову лише у випадку періодичних сигналів, для яких частоти коливань однакові або кратні. Але на практиці використовують поняття «миттєвої фази» сигналу, яка змінюється в часі.

Якщо маємо два сигнали $x(t)$ та $y(t)$ з миттєвими фазами $\varphi_1(t)$ та $\varphi_2(t)$, фазова синхронізація між ними означає виконання такої умови:

$$|n\varphi_1(t) - m\varphi_2(t)| \approx const$$

де m та n – цілі числа, що показують кратність синхронізації.

Чисельною характеристикою фазової синхронізації є *коефіцієнт фазової синхронізації* (phase locking value, PLV), який описує рівень синхронності зміни фаз та стабільність різниці фаз для двох сигналів в певному частотному діапазоні. КФС можна розрахувати за формулою:

$$PLV(f, t) = \frac{1}{N} \sum_{n=1}^N |e^{j\theta_n(f, t)}|,$$

де N – кількість вимірювань сигналу, які використовуються;

f – частота, на якій досліджується синхронізація;

t – момент часу, для якого розраховують миттєві фази;

$\theta_n(f, t) = \phi_{1n}(f, t) - \phi_{2n}(f, t)$ – різниця фаз для n -ої пари сигналів на частоті f

в момент часу t .

Якщо неможливо провести багато вимірювань для пар сигналів, то використовується інший коефіцієнт, що оцінює стабільність різниці фаз протягом певного часового вікна:

$$SPLV(f, t) = \left| \frac{1}{\delta} \int_{t-\frac{\delta}{2}}^{t+\frac{\delta}{2}} e^{j(\phi_1(f, \tau) - \phi_2(f, \tau))} d\tau \right|,$$

де δ – тривалість часового вікна.

Значення коефіцієнтів, що характеризують фазову синхронізацію, лежить в межах між 0 та 1; якщо PLV або SPLV близьке до 1, то це говорить про високий рівень синхронізації між двома сигналами.

В загальному випадку розраховувати характеристики синхронізації треба для коливань на одній частоті, але на практиці вимагається, щоб обидва сигнали мали вузькосмуговий спектр. Для цього сигнали попередньо фільтруються смуговими фільтрами. Також необхідним є визначення миттєвої фази для кожного з двох сигналів в певні моменти часу. Це найчастіше виконується з допомогою перетворення Гільберта або вейвлет-перетворення.

Взаємна інформація

Взаємна інформація – це міра обсягу інформації, яку містить один випадковий процес про інший випадковий процес. Взаємна інформація між двома сигналами може бути розрахована лише у випадку, коли вважається, що ці сигнали – випадкові і зв'язані один з одним. Іншими словами, величина взаємної інформації – це те, наскільки зменшиться незнання характеристик одного процесу, якщо ми знаємо характеристики іншого процесу, що пов'язаний з першим. Взаємна інформація між випадковими процесами X та Y може бути розрахована за виразом:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right),$$

де $p(x, y)$ – взаємна густина розподілу імовірності випадкових процесів X та Y ;

$p(x)$ та $p(y)$ – їх густини розподілу імовірностей, відповідно.

Основною проблемою при розрахунку взаємної інформації є необхідність мати густини розподілу для сигналів, які аналізуються. Для цього існують різні методи оцінки густин розподілу, зокрема, на основі гістограм. Також необхідно проводити статистичні тести для того, щоб виявити, чи дійсно сигнали пов'язані, чи ненульове значення взаємної інформації отримана випадково і не має смислу.

8.4. Підсумки до розділу 8 та рекомендована література

Всі методи аналізу сигналів було класифіковано на лінійні та нелінійні методи відповідно до того, як пов'язана величина, що характеризує сигнал, зі значенням величини сигналу. Також було виділено уні-, бі- та мультиваріативні методи, які відрізняються тим, скільки каналів сигналів одночасно використовується при аналізі. Ці класифікації проілюстровані найбільш розповсюдженими методами уні- та біваріативного аналізу. Розглянуто ентропію перестановок, аналіз рекурентності, крос-спектральний аналіз, аналіз взаємної інформації та фазової синхронізації. Нажаль, майже відсутні доступні підручники з нелінійного аналізу сигналів, тому основні відомості щодо цих методів та їх застосування треба шукати в наукових журналах.

8.5. Питання для самоперевірки

1. Опишіть відмінності між лінійними та нелінійними методами аналізу сигналів. Опишіть відмінності між уні-, бі- та мультиваріативними методами аналізу сигналів.
2. Наведіть приклад та опишіть сутність спектрального аналізу за Фур'є та вейвлет аналізу.
3. Наведіть приклад та опишіть сутність аналізу когерентності.
4. Наведіть приклад та опишіть сутність аналізу ентропії перестановок.
5. Наведіть приклад та опишіть сутність аналізу рекурентності.
6. Наведіть приклад та опишіть сутність кросс-спектрального аналізу та аналізу когерентності.
7. Наведіть приклад та опишіть сутність аналізу взаємної інформації.
8. Наведіть приклад та опишіть сутність аналізу фазової синхронізації.

ВИКОРИСТАНА ЛІТЕРАТУРА

1. Френкс Л. Теория сигналов. – М.: Сов. радио, 1974. – 344 с.
2. Оппенгейм А., Шафер Р. Цифровая обработка сигналов. – М.: Техносфера, 2006. – 856 с.
3. Айфичер Э.С., Джервис Б.У. Цифровая обработка сигналов: практический подход. – М.: Издательский дом «Вильямс», 2004. – 992 с.
4. Марпл-мл. С.Л. Цифровой спектральный анализ и его приложения: Пер. с англ. - М.: Мир, 1990. - 548 с.
5. Робинсон Э.А. История развития теории спектрального оценивания. / ТИИЭР. – 1982. – т.70, № 9. – С. 6 – 32.
6. Лайонс Р. Цифровая обработка сигналов. – М.: ООО «Бином-пресс», 2006. – 656 с.
7. Сергиенко А.Б. Цифровая обработка сигналов: Учебник для вузов. 2-е изд. - СПб.: Питер, 2007. - 751 с.
8. Бахтиаров Г.Д., Малинин В.В., Школин В.П. Аналого-цифровые преобразователи. – М.: Советское радио, 1980. – 280 с.
9. Алгоритмы и техника цифровых измерений: Учеб. пособие / Е.Ф.Клещевников – К.: УМК ВО, 1988. – 108 с.
10. Воробьев Н.Н. Теория рядов. - М.: Наука, 1979. - 408 с.
11. Ильин В.А., Позняк Э.Г. Линейная алгебра - М.: Наука, 1974. - 296 с.
12. Гоноровский И.С. Радиотехнические цепи и сигналы. – М.: Радио и связь, 1986. – 512 с.
13. Ганеев Р.М. Математические модели в задачах обработки сигналов. – М.: Горячая линия-Телеком, 2004. – 80 с.
14. Фрейзер М. Введение в вейвлеты в свете линейной алгебры. – М.: БИНОМ. Лаборатория знаний, 2008. – 487 с.
15. Ахмет Н., Рао К.Р. Ортогональные преобразования при обработке цифровых сигналов: Пер. с англ./ Под ред. И.Б. Фоменко. – М.: Радио и связь, 1980. – 248 с.
16. Mallat S.G., Zhang Zh. Matching pursuits with time-frequency dictionaries. // IEEE Transactions on signal processing. – 1993. – vol. 41, № 12. – P. 3397 – 3415.
17. Chapa J.O., Rao R.M. Algorithms for designing wavelets to match a specified signal // IEEE Transactions on signal processing. – 2000. – Vol. 48, №12. – P. 3395–3406.
18. Chen S.S., Donoho D.L., Saunders M.A. Atomic decomposition by basis pursuit // SIAM Review. – 2001. – Vol. 43, №129. – P. 1–29.
19. Bronzino J. D. The biomedical engineering handbook. Boca Raton: CRC Press LLC, 2000.

20. Малла С. Вейвлеты в обработке сигналов: Пер. с англ. – М.: Мир, 2005. – 671 с.
21. Харкевич А.А. Спектры и анализ. - М.: Государственное издательство физико-математической литературы, 1962. - 236 с.
22. Голд Б., Рэйдер Ч. Цифровая обработка сигналов. – М.: Сов. радио, 1973. – 367 с.
23. Гольдман С. Гармонический анализ, модуляция и шумы. – М.: Издательство иностранной литературы, 1951. – 422 с.
24. Harris, F.J. On the use of windows for harmonic analysis with the discrete Fourier transform / F.J. Harris // Proceedings of IEEE. – vol. 66, № 1. – P. 51 – 83.
25. Штарк Г.-Г. Применение вейвлетов для ЦОС. – М.: Техносфера, 2007. – 192 с.
26. Чуи К. Введение в вейвлеты. – М.: Мир, 2001. – 412 с.
27. Блаттер К. Вейвлет-анализ. Основы теории. – М.: Техносфера, 2004. – 280 с.
28. Переберин А.В. О систематизации вейвлет-преобразований // Вычислительная математика и программирование. – 2001. – т.2 – С. 15 – 40.
29. Daubechies I. The wavelet transform, time-frequency localization and signal analysis // IEEE Transactions on information theory. – 1990. – Vol. 36, № 5. – P. 961 – 1005.
30. Брейсуэлл Р. Преобразование Хартли: Пер. с англ. – М.: Мир, 1990. – 175 с.
31. Хемминг Р.В. Цифровые фильтры. - М.: Советское радио, 1980. - 224 с.
32. Богнер, Р. Введение в цифровую фильтрацию / Р. Богнер, А. Константи́дис. М. : Мир, 1976. – 216 с.
33. Андерсон Т. Статистический анализ временных рядов. - М.: Мир, 1976. - 755 с.
34. Свешников, А.А. Прикладные методы теории случайных функций. – М.: Наука, 1968. – 463 с.
35. Гонсалес Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005. – 1072 с.
36. Гонсалес, Р. Цифровая обработка изображений в среде MATLAB / Р. Гонсалес, Р. Вудс, С. Эддинс. – М. : Техносфера, 2006. – 616 с.
37. Ту, Дж. Принципы распознавания образов / Дж. Ту, Р. Гонсалес. – М. : Мир, 1978. – 411 с.
38. Дуда, Р. Распознавание образов и анализ сцен / Р. Дуда, П. Харт. – М. : Мир, 1976. – 507 с.
39. Теория автоматического управления: Учеб. для вузов по спец. «Автоматика и телемеханика». В 2-х ч. Ч. II. Теория нелинейных и специальных систем автоматического управления. / А. А. Воронов, Д. П.

Ким, В. М. Лохин и др.; Под ред. А. А. Воронова.— 2-е изд., перераб. и доп. — М.: Высш. шк., 1986.— 504 с.